

STIC Search Report

Biotech-Chem Library



STIC Database Tracking Number 193541

TO: Nancy Vogel
Art Unit: 1636
Location: REM-2A70&2C70
Serial Number: 10/689200
Wednesday, May 25, 2005

From: Beverly Shears
Location: Biotech-Chem Library
REM 1A54
Phone: 571-272-2528
beverly.shears@uspto.gov

Search Notes

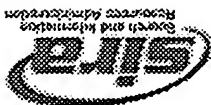
Protein Sequence Searches – February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension .rnp) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (uniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.



This Page Blank (uspto)

STIC-Biotech/ChemLib

153541

From:
Sent:
To:

Vogel, Nancy
Monday, May 16, 2005 2:46 PM
STIC-Biotech/ChemLib

Subject:

sequence search 10/689,200

Please do a sequence search of SEQ ID NO: 1 and 2

and return results to me on paper asap.

Thanks,

Nancy Vogel
Patent Examiner
Art Unit 1636
Office: Remson 2A70
Mail Box: Remson 2C70
(571) 272-0780

STAFF USE ONLY

Searcher:
Searcher Phone: 2-
Date Searcher Picked up:
Date Completed:
Searcher Prep/Rev. Time:
Online Time:

Type of Search
NA#: _____
Inference: _____
SPDI: _____
S/L: _____
Oligomer: _____
Encode/Transl: _____
Structure#: _____
Text: _____
Invention: _____
Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

Date completed: _____
Searcher: Beverly e 2528
Terminal time: _____
Elapsed time: _____
PU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
STIC
CM-1
Pre-S
Type of Search
N.A. Sequence
A.A. Sequence
Structure
Bibliographic

Vendors
IG
STN
Dialog
APS
GenInfo
SDC
DARC/Questel
Other CCN

This Page Blank (uspto)

Vogel, N.
107689200 Page 1
Seq. IDs 172

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2005, 13:18:48 ; Search time 6175 Seconds
(without alignments)
10923.022 Million cell updates/sec

Title: US-10-689-200-1

Perfect score: 1392

Sequence: 1 acgaacaccatcatagatc.....ccggcggaagtcgtgat 1392

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 4708233 segs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

GenBml:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_com:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_str:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	628.8	45.2	11493	1 AF283498	AF283498 Azospirillum
C 2	555.6	39.9	300883	1 AE016809	AE016809 Vibrio vu
C 3	541.2	38.9	248880	1 AP005348	AP005348 Vibrio vu
C 4	516.2	37.1	302100	1 AP005087	AP005087 Vibrio pa
C 5	431	31.0	311050	1 BX294133	BX294133 Pirellula
C 6	171.2	12.3	110000	1 AE017282_06	Continuation (7) of
C 7	133	9.6	10676	1 AE008293	AE008293 Agrobacte
C 8	133	9.6	10707	1 AE009321	AE009321 Agrobacte
C 9	122.8	8.8	348077	1 AP003000	AP003000 Mesorhizo
C 10	119.8	8.6	300150	1 AP005945	AP005945 Bradyrhiz
C 11	118.2	8.5	110000	1 RME591985_07	Continuation (8) of
C 12	116	8.3	196050	1 AL646058	AL646058 Ralstonia
C 13	113.2	8.1	298700	1 AP005951	AP005951 Bradyrhiz
C 14	109.6	7.9	12129	1 AE007223	AE007223 Sinorhizo
C 15	104.8	7.5	299750	1 AP006575	AP006575 Gloeobact
C 16	104.4	7.5	339650	1 AP005583	AP005583 Nostoc sp
C 17	100.6	7.2	348068	1 BX572604	BX572604 Rhodosphe
C 18	98.4	7.1	300700	1 AP006573	AP006573 Gloeobact
C 19	98.2	7.1	110000	1 BX950851_33	Continuation (34) o

20	95.2	6.8	17155	1 AY273169	AY273169 Rhodobact
21	95	6.8	4148	1 KPNNA5FEC	KPNNA5FEC Klebsiella
22	93.4	6.7	135551	1 D90906	D90906 Synechocyst
23	89.8	6.5	3253	1 SYONIRNRT	SYONIRNRT D12723 Synechococ
24	89.2	6.4	300363	1 AE016781	AE016781 Pseudomon
25	87	6.2	10798	1 AE004604	AE004604 Pseudomon
26	86	6.2	300300	1 AP005373	AP005373 Thermosyn
27	84.6	6.1	349142	1 BX572539	BX572539 Rhodosphe
28	82.4	5.9	299650	1 AP005955	AP005955 Bradyrhiz
29	82.4	5.9	1371	6 AR386463	AR386463 Sequence
30	80.2	5.8	4080	1 SYOCMP	SYOCMP D26358 Synechococ
31	79.4	5.7	298750	1 AP005375	AP005375 Thermosyn
32	77.8	5.6	302550	1 AP006581	AP006581 Gloeobact
33	75.2	5.4	11318	1 AE008005	AE008005 Agrobacte
34	75.2	5.4	11330	1 AE009039	AE009039 Agrobacte
35	74.6	5.4	6449	1 PLNRTABC	PLNRTABC P19598 P. lammosum
36	74.4	5.3	347400	1 AP003591	AP003591 Nostoc sp
37	73.8	5.3	7520	1 AE005736	AE005736 Caulobact
38	72	5.2	132106	1 SYCSLH	SYCSLH D64006 Synechocyst
39	71.8	5.2	310325	1 AE016864	AE016864 Pseudomon
40	70	5.0	3218	1 AE001333	AE001333 Synechoco
41	69.2	5.0	11321	1 AE005737	AE005737 Caulobact
42	65.8	4.7	3719	1 AY283674	AY283674 Cyanobact
43	63.4	4.6	2210	1 AYNA5ST	AYNA5ST X83602 A. vinelandi
44	63.4	4.6	4080	1 SPCNRT	SPCNRT X61625 Synechococ
45	60.8	4.4	110000	1 CP000011_11	Continuation (12) o

ALIGNMENTS

RESULT 1	AF283498	11493 bp	DNA	linear	BCT 02-JUL-2002
LOCUS	Azospirillum brasilense nitrate assimilation gene cluster, complete sequence.				
DEFINITION	AF283498.1 GI:21666012				
ACCESSION	AF283498				
VERSION	AF283498.1				
KEYWORDS	Azospirillum brasilense				
SOURCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales; Rhodospirillaceae; Azospirillum.				
ORGANISM	1 (bases 1 to 11493)				
REFERENCE	Steenhoudt, O., Placek, D., Verreth, C. and Vanderleyden, J.				
TITLE	Nitrate assimilation in Azospirillum brasilense Sp245: mutational and expression analysis of the nas gene cluster				
JOURNAL	Unpublished				
AUTHORS	2 (bases 1 to 11493)				
TITLE	Direct Submission				
JOURNAL	Submitted (29-JUN-2000) F.A. Janssens Laboratory of Genetics, Catholic University of Leuven, Kard. Mercierlaan 92, Heverlee 3001, Belgium				
FEATURES	Location/Qualifiers				
source	1..11493				
	/organism="Azospirillum brasilense"				
	/mol_type="genomic DNA"				
	/strain="Sp245"				
	/db_xref="taxon:192"				
gene	/complement (526..2190)				
	/gene="nasD"				
CDS	/complement (526..2190)				
	/gene="nasD"				
	/codon_start=1				
	/transl_table=11				
	/product="nitrate transporter Arpase"				
	/protein_id="FAM73541.1"				
	/db_xref="GI:21666013"				
	/translation="MALIKNVAKSGGVNVRGDIATAGEFVAIVGSGAKTT LVMKGLAPDAGIVLFGAPVKGAEKGVFQNSLMPHLSVDNALAYDAHK GKSPRGVITRAAVTGVGATDRPELSGKMQRFAGFALMSPEMLLDEL SALDALTRAKLODEIEAIWRKRTKTVLLITNDVDAIILLADRILPLPGAVLGPAP				

Db 4250 GCCCAAGAGCTGTGGGCGCAAGATCCAGGCGGAGCGCTGCTCCGTCAACCCCGCCG 4191
 Qy 666 GCAATGCGCGTCCAGCTGAGAGCCGCGACCATTTTGGTTATTCGTGCGGAGCGCGT 725
 Db 4190 GCAGATGCGCGCAGCGCTGAGAGCCGCGGACCATTTTGGTTATTCGTGCGGAGCGCGT 4131
 Qy 726 GAAACGAGCGCGTGTTCAGAGGCGATAGCGGTGCGGTATCAACCGATGAAGACTCTG 785
 Db 4130 GAACCGAGCGCGGTATGAGAGGCGATGCGCGGTGCGGTATCAACCGAGATCTG 4071
 Qy 786 GAAGGACGCGCGGAAAAAGTCTTTCGCGGTGAGCAAAACATTTGGGCGGAGAAATACCCCA 845
 Db 4070 GAAGAAACACCCGAGAGAGGTCTTCGCGGTACCGAGGCTGGGCGCGCAAGAACCCCA 4011
 Qy 846 CACTTATCTGCGCGTGAACCAAGCGCTGATTCGCGCGCGCATCTGCTGAGCGCGCA 905
 Db 4010 GACCCATCTGCGCGCTGGTCAAGCGCTGATTCGCGCGCGCATGCTGCTGAGCAAGACG 3951
 Qy 906 TAAACAAAGACCGCAAGAGCCATCGAATGCTGCGCGCAAAAACATACGTCGCTGTA 965
 Db 3950 CAACGCAACCGCGCGAGCGCGGTGAAGATCTGCGCAAGTGGAGTATGCGGCGCGCA 3891
 Qy 966 CGTGAAGTCTGCGCGGTAGCAAGACCGGACCTTTCGATTAAGAAAGACGATAACG 1025
 Db 3890 CGCAACGCTATCGCACTCGATGACCGGACCTTTCGATTAAGAAAGGCGCAACGCG 3831
 Qy 1026 CGCGCTACCGGACTTCAACACTTCTTTCGCGCACGCGCGCGCATCTGCTTACAGCAG 1085
 Db 3830 GGGGGTTCGCGACTTCAAGCTCTTCTTCGCGCTACAAAGCCACTATCCCTTCTACTCGA 3771
 Qy 1086 TCGAGTCTGATATCTGACCAAGCTGAGCGCTGCGGCGCATATGATGAATTCAAACCGGA 1145
 Db 3770 CGCGCTGCTGATCTGACGAGATGCGCGCTGCGGCGCATATGCGGCGCAAGCGCGA 3711
 Qy 1146 CAACGTGATCTGATATACCGGCAAGACGCTTACCGCGCGGACATATTCGCGCGCGG 1205
 Db 3710 CGCGCTGATGACGAGACCGCGCAAGGCTTACAGCGGAGATCTTACCTGAAGCGCGC 3651
 Qy 1206 CAAGAACTGCTGCGAGAGGCAAGCGCAAGCGCGCAAGCTTCCC--TSCCGATACCTC 1262
 Db 3650 GCGGCTGCTGCTGAGAGAGGCAAGCGCGCAAGCGCGCAAGCTTCCCCTCGAGACGACG 3591
 Qy 1263 GATCAAGCGCTGCGAGATTTCTTCAATCGAACAAAGTCCGCTTCGATGCCAAAGCCCA 1322
 Db 3590 CTACAAGCGCGTGAATACGATTCATCGACGCGCATCGCTTACGACGCGCGCAAGCCCA 3531
 Qy 1323 CGATTTCCTCGCAATTTGCGATAGCTCTGAAGAGCAAGCAACCGTACCGCGCGCA 1382
 Db 3530 CGAGTACTGACGACGACTGCGCTCGAAGGCGCGGCGAGCGGTGCGCA 3471
 Qy 1383 GGTCTGG 1390
 Db 3470 ACTGCTG 3463

RESULT 2
 AE016809/c 300883 bp DNA linear BCT 29-SEP-2003
 LOCUS Vibrio vulnificus CMCP6 chromosome II section 2 of 6 of the
 DEFINITION Complete sequence.
 ACCESSION AE016809 AE016796
 .VERSION AE016809.1 GI:27358288
 .KEYWORDS
 SOURCE
 ORGANISM

Vibrio vulnificus CMCP6
 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 Vibrionaceae; Vibrio.

REFERENCE

AUTHORS

Kim, Y.R., Lee, S.E., Kim, C.M., Kim, S.Y., Shin, E.K., Shin, D.H.,
 Chung, S.S., Choy, H.B., Progulski-Fox, A., Hillman, J.D., Handfield, M.
 and Rhee, J.H.
 Characterization and Pathogenic Significance of Vibrio vulnificus
 Antigenes Preferentially Expressed in Septicemic Patients

JOURNAL Infect. Immun. 71 (10): 5461-5471 (2003)
 PUBMED 14500463
 REFERENCE 2 (bases 1 to 300883)
 AUTHORS Jeong, H., Moon, Y.H. and Kim, J.U.
 TITLE Direct Submission
 JOURNAL Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeommin-dong,
 Yuseong-gu, Daejeon 305-811, South Korea
 REFERENCE 3 (bases 1 to 300883)
 AUTHORS Rhee, J.H., Kim, S.Y., Chung, S.S., Lee, S.E. and Choy, H.B.
 TITLE Direct Submission
 JOURNAL Submitted (13-DEC-2002) Department of Microbiology, Genome Research
 Center for Enteropathogenic Bacteria, Chonnam National University
 Medical School, Hak-1-Dong, Dong-gu, Kwang-Ju 501-746, South Korea

FEATURES

source

gene

CDS

gene

CDS

Gene
CDS

/locus_tag="VV20275"
 565..2553
 /locus_tag="VV20275"
 /note="COG2199"
 /codon_start=1
 /transl_table=11
 /product="F03: GDSF domain"
 /protein_id="AA007240.1"
 /db_xref="GI:27358290"
 /translation="MCSLSLSLTKNLLPILFTLLIAGGHHFMSYEKQDQLO
 NHRISLAIAGAOONLQASVLFNDQKMOEILSFSDPSILIVELSLLEGVAVR
 BHPVVDNHLVLPYPERFTFYSORIDVEONAIATLSTASTQLEOMVTRLSSV
 VPLAILLSGSTLYLIVOKVVIDPVYKLNESMODIVEGRIKGSLSHSDENGELVN
 ARANSQRLALRDSOVHTLDRLEHKAATDVITAEVQTRLLVYNEHGDILHNAAR
 MLFQSDAQRILAQRKSELMCKEFSVANALTSKNLEKLVLCRPLTSGSKLTKIST
 RLFSKRRAFLFSIODVTEIETALNRORLAGVFNESQDGLVVDHNGLLTMTPAVTR
 MGYPOEILIGKSLKAFPMOQIOTIMPTIOESLBEHYGOMQGEVLKHYGNLIPMA
 KVSRLSKVGVQVEFPMVILISDLSGVKEMERLEYLAHHDALGLANRAKGLRHTLDE
 LKLCNTHSEFPALVYMDLDGFKTINTPGRHDADHYLAKVSRRLANVNSGDIYVALS
 GDFVLLIRADNHSVTLLSERLSELRPIETLNTQTVQVGSIGVXLVGPQEKMER
 ILKSADTANYKAKSGGQAILIGSEVGACLSDVTS"
 complement (2838..4703)
 /locus_tag="VV20276"
 complement (2838..4703)
 /locus_tag="VV20276"
 /note="similar to mostly Fe transport; COG1629"
 /codon_start=1
 /transl_table=11
 /product="Outer membrane receptor protein"
 /protein_id="AA007241.1"
 /db_xref="GI:27358291"
 /translation="MEGRVKLLVNGASQPCVFDGCPYFINSANVSVDPMKSYEI
 VKGASSLHGSDAIIGGVAFETKDPADPLFKNGVFGQAKLSYSSSDKSPSERVAAK
 RPDNVLAVALTRDDEKELNFSKAPDYVESQDYVYNDLLIKLQSLQSLDRHLEF
 LGEVYNQDSDIASSSYKNFNSDSTTKONRLGAKHILWLDITITDSITSRLLTWLDE
 ENGLTRFKBASGVPVPNGDNOQKQDOYAEKRIELETFQDFDELTHYLVVGLS
 YKTSRISSNTRFENSDPNTPDKTYVTVQKQDOYAEKRIELETFQDFDELTHYLVVGLS
 DYFSTDPSTVDSEPFKESDSAFTGALGATTKYVAGQGITLFTQISQFRAPSRRELTYT
 YNPGHGYNRNPNLESKSLSYELGVHNTASTSTALYSYDVTDEIQTSTNG
 GLETVYVNISEATIKIGIEVSNTLDLNRITANLPLGWTTRFVAAYTVGDEBGNPLNSV

NPENAVGVVYDAPNTLMGTSIKVNTYAKNSGHINLDGNNGGTANQVLEPATVIDL
TAYKPPQDVTVAAGIMNLTNEKXYLMNDIRGEKEISRLDTQAEFRYNTISVKYEF"
gene
5180. .6094
/locus_tag="VV20277"
CDS
5180. .6094
/locus_tag="VV20277"
/note="COG0583"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator"
/protein_id="AA007242.1"
/db_xref="GI:27358292"
/translation="MILITKGVESQSPITLALHILDAIDRSGSPASANEILRAPS
LSYQIQEDODLIMIPDRSGHAKNTPDARLILBERGRTILATETCUNDATITLANG
EDLITAYDGIIGIHFFSLVDALGVNSKTRVLRQEBITLAGEWSLQSGADILVCPK
LEVLPHVDKADTIGKMSMTVVAAPNHYNRSGEFDYAEKRIIADTAEGOPAI
TVNIIOKQPLRTVTLNQAKVDALVTGIGTGLPTDIATPLIEAGKLEIEGAQOHLE
IILAKRNQMGKASWCICYLKKQMKLK"
gene
complement(6143. .8821)
/locus_tag="VV20278"
CDS
complement(6143. .8821)
/locus_tag="VV20278"
/note="other COG candidate exists; similar to
acetyltransferase; COG0454"
/codon_start=1
/transl_table=11
/product="Histone acetyltransferase HPA2"
/protein_id="AA007243.1"
/db_xref="GI:27358293"
/translation="MFKLOKLFKPTSIIVAGASORHLSAGNIYMKNLDSGPDGTTIP
VMPKSVAGVAVPSVTELPVPIAIILCTRERVALFEKOLAEIGTEFVIVLADIT
DYHODQGVADACLRARBSNRVUGRNSGILPQKENSFSPYTKRGTIAIS
OSARVCTTLIDRANKKEIGSAFVSTGNGLDIDFELDLSTDSKTDAILLYDSIT
DARFTSAARAAARNRRIIVLKGRSILGGRALNKSADUTLIDVDSAIRSGMLVN
NHLEPAAVETLHVSPLRGERLAIITNGAPALMAVDTLFERGQLANISSETIEM
SOHLPSWSHNNPIDIVGADKGRVATTINALIDSDCADVLIMHSPANSDETEAO
AIVDAIOSHPHRRFNILITWMSGQOTAKPARMLITKAGIPTYRTPESAVVAFWHILEY
RNQOKMETPTTAPLIVSDVNAAKAVSEHLMDEKIAQIDTHQGLSLAKLPHVPL
PRTIADTSEAVHIAEOIGTPYAVVKLRSPDIANKSVQGMNLNSQEVASADAIL
DRKLAVPSANIHGLVQVQAKLAGSEERVKVITTKVGPVILLQSGSEWSEADA
LDLPLPLNLTAVQIVRAIKEGHILQKLPEDVYGLPAKFLVRIQSVVLELPQKE
LDHPLVANGHEFTILDADLTLENYLGHQGERLAIPRPPEVETWLOKGEIILRP
LIDEPESHAEFTINNSKEDLYKRPSPDSGEFHEMLANLTOIDYRBAFVAVSS
GEAKTIGVARALINPENTAEFPAILIRSLDKGKGLKVTLMNKIIDICRQGTAKMTGM
TPTNRGMLTLAQKGLSLDIOPEGDVADVPLNP"
gene
9046. .9621
/locus_tag="VV20279"
CDS
9046. .9621
/locus_tag="VV20279"
/note="COG3266"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein conserved in bacteria"
/protein_id="AA007244.1"
/db_xref="GI:27358294"
/translation="MKKVAIITLSTVLSACASDNTYANTTSHREEFKIEPIPEML
ABEIVQEVSTIQQVVMKEPQTEKRVVAKLIPSESTTKVAVKLVPSKQODKQRGY
TYQVAVGQAKVDYFASKLPDQGPIMENYKLVNGTKWFTVLYGDFATSTAKKAIA
QJPOEFKELOPFKSIDAIKNSHYPTLNKLN"
gene
9788. .10783
/locus_tag="VV20280"
CDS
9788. .10783
/locus_tag="VV20280"
/note="similar to ATP-grasp enzymes; COG1181"
/codon_start=1
/transl_table=11
/product="D-alanine-D-alanine ligase"
/protein_id="AA007245.1"
/db_xref="GI:27358295"
/translation="MNRNVLLCGGSSSEHISLSSEVYLQOGLIENNVYLKYEIK
NGMFQOKERLVYLDIHTKSVSDENFBSIHIDFIVPCIHGFGEYGDIOCSPELAGI
PLTGGPEASNSFNKTSKLVMDAIGINTPLPLTENNESHOISRAEPQKGLTF
VAAAGSSVGCSTYKIEQLSDAIDKAGFSGHQLVVEKAVKPRELEVASAYEMNGQIH

ISKPRITAPDGAFFSYDEKYSAGSHSITTEVAKNLTLEQOLATITOLCSKRVFROKMLR
HLSRIDPFTLSBEIEILNEVNTFPQGTTKISMEPKKLQHNHKEHETLADICERSL"
gene
11080. .11454
/locus_tag="VV20281"
CDS
11080. .11454
/locus_tag="VV20281"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="AA007246.1"
/db_xref="GI:27358296"
/translation="MAISHYRGFNLOSAFNDNTIMOVKIKRVLQSLSAVKKSIDMW
COTASIIDPERSSLDRKQEAATGPAENFNGPAILKNDGEPNAWCFPNGKLIKQSK
MAIQRTIEAYLILAKQAEQOK"
gene
11457. .11768
/locus_tag="VV20282"
CDS
11457. .11768
/locus_tag="VV20282"
/note="eIF-1/STU1; COG0023"
/codon_start=1
/transl_table=11
/product="Translation initiation factor 1 related protein"
/protein_id="AA007247.1"
/db_xref="GI:27358297"
/translation="MTLVSTERTGRIRKPEPRKASRPKGVIRIOPETGRKGVGV
VTGLDDDAPLKLIILAEELKIKCGGGSVQDGINIEIGDARDKIKTHLEKKGKVKVQAG
G"
gene
12360. .12782
/locus_tag="VV20283"
CDS
12360. .12782
/locus_tag="VV20283"
/note="COG0499"
/codon_start=1
/transl_table=11
/product="S-adenosylhomocysteine hydrolyase"
/protein_id="AA007248.1"
/db_xref="GI:27358298"
/translation="HDFEFNVLFRVFPSTIRAGAEYFRHVQATYVRMLSGSLPIPNM
Query Match 39.9%; Score 555.6; DB 1; Length 300883;
Best Local Similarity 64.9%; Pred. No. 9,6e-97;
Matches 876; Conservative 0; Mismatches 459; Indels 15; Gaps 3;
25 TCGAAGAAATTTGTTATGACGTTATCGGCTTGCTGCTAGCCGTTGGGCTTGACAGATTGG 84
118404 TCGATGAAAAAGAGATGACAGTGTGCGGTATTAAACAAGTTGGTGGCGTTTGCTTACCG 118345
85 CCGGATGTCGGGGCAGTTGCGCAAGCTGGAAGAAAGAAAGATTGGAATTCGCGCTTCATCAAG 144
118344 GTTGTACTGCGAAGATCGGCGAGGCTGGAAGAAAGAAAGAAAGATTGGAATTCATCAAA 118285
145 CTCACCGACATGCGCGCGCTGCGGTGGCGCGCGCGGAAAGAGCTTCTTCAGAGACGAGGCG 204
118284 TTGACCGACATGCGACCGTTGGCGAGTCGTTTGAAGAAAGATTTTGTAGAGATGAAGGG 118225
205 CTGTTGTCGCAACTGGAACCGCAAGCCCAACTGGAAGGTGTGATGATAGGCTCGTAAT 264
118224 CTGTATGTACGCTGAAGAACGAGGCGCAACTGGAAGGTCTGCTCATGCGCGTGAATGAT 118165
265 GGGGAACGTGAAGCGCTGCGACATGCTGTGGCGCGGCGCGTGAAGCGGCGAGCGTGGCTTC 324
118164 GGTGAACGTGAAGCGCGCGACATGCTGTGGAGGCGCAACACTGGGTGCCACATCGGCATT 118105
325 GGCACCAAGCGCGATTCAGAGGTGCGCTTCAGACATGGCTTCAGACGCGACCGATTAAG 384
118104 GGCACCGCAAGCCAAAGTATACACCGCATTCAGATGATGATCTCAATGTGTAACCGCATCAC 118045
385 GTGTCCATGAATTCGCGATCAATGAAGCCGAATACCGCTGGA---AGCGGTAAA 441
118044 GTGTCCAGAGATTTTGGCGAGCAAAATGAAGCCCAATTGTGTAAGAACAGTACGCGTAAA 117985
442 CCGGTGCATCCGATCAAGGCAATTAATCTAAACCGGTGTGGAAGAAATACAAAGCCGAA 501

Db	117984	CCCGTTCA	CCCCATGA	AGCGGATGCCCTTAA	ACGTGTGGTAA	CAAGTTAC	CCGCGACGAA	1179292
QY	502	GGCAAGCCGTT	CAATATATGG	CGATGACCTT	CCCGG	CCGGATCT	CAACAACATCAAACTGCGT	561
Db	117924	GGCAAAAGCGTT	CAATATATGG	CGATGCGTTT	CCCCGTTT	CAACGATAACT	ATCAAGATCGCGC	1178655
QY	562	TACTGCGCTGG	GGGCGTGGGTAT	TCATCCGTGCTACT	ATATTCGGCGCGCA	---GACATT	618	
Db	117864	TACTGCGCTGG	GGGCGGATTCAT	CTGGCTTTTATATG	CCGCGATAAAGCGACAA	C	1178059	
QY	619	TCGCGCCAAAT	TCGCGCAGACCG	CTTGTGTGCGATGAC	CCCCCGCGCCGCAAAATGCCGTCC		678	
Db	117804	AGTGGCCAAAT	TATATATGCG	ACGTGCGTCAAGGTG	ACCGCGACCA	CAAAATGCCAGCC	1177454	
QY	679	ACCGTGAAG	CCGCGCAACATTTTCG	GTATATGCGTCCGAGCGCTG	GAAACCAAGCC		738	
Db	117744	ACCATGGAAG	CGGCGGCACCATCAAA	GGCTATATTCGTTGTG	GAACCGTGGAAACCAAGCCG		1176859	
QY	739	GTGTTCAAGG	GGCATATGAGGCGTGC	CGGTGATCA	CGATGAAAGAACT	CTGGAAAGGACAGCCG	798	
Db	117684	GTGTTTAAAGG	CGCATTTGTGTGCCCCG	GTGACCGCATATG	ATGATCTGAAAAACAACCA	CCCA	1176235	
QY	799	GAAAAAGT	CTTCGCGCGTGACCAAA	CAATGGGCGGAGAAATAT	CCCCCAACCTTATCTGCGC	858		
Db	117624	GAGAAAGTATTT	GGCGTTGCGCAAA	ATATGGGCGGAAAAAT	ATCCCAATATCCCATATCCG	1175655		
QY	859	GTGACCAAG	CGCGCTGATTTCCGCGCGCGATCT	GTGCTGGA	CGCGCAATATCAAGAACCG	918		
Db	117564	GTGCTCAAG	CTCTGATCCGTGCGCGCGCATTTG	GTATGATGA	AAACGATTAACCGCATTCGC	1175059		
QY	919	AAGGAAGCAT	TCGAATATGTGCGCGCAAAA	CAATAGCTGCGTCTGA	CTGTGAAGTGTCTG	978		
Db	117504	GCCGAAGCG	GTGAACCTGCTTCA	CGCACGACGATAGTGGGCGCGCATG	CGGAATGTGATC	1174454		
QY	979	GCGGCTACAT	GAAGCGGCACTTTCGAATAT	CGAAAAAGAGATTAAG	CGCGCGTACCGGAC	1038		
Db	117444	GCCATTTCC	ATATGATCTGCGACGCTTTG	ATATACGAAAAAGCGATATAGGCA	CAATGTCGCCGAT	1173859		
QY	1039	TTCAACACCTT	CTTTCGCGACCGGCGCCAG	CTATTCGTTCT	CAAGCAGTGCATCTG	1098		
Db	117384	TTTATATG	TTGTTCTTTCGCTAC	AAACGACCTATCCCTTA	CTACAGCGACGGAATTTG	1173259		
QY	1099	CTGACCC	AGCTGAAGCGCTGCGGCGCAT	GTCAATGAATTCAA	ACCGGACCACTGTGATCTG	1158		
Db	117324	CTGACCAAAAT	ATGCGCGTTGGGGA	CAAAATCGAAATCAG	AAACCGGACCTCATGTGTCAATG	1172655		
QY	1159	GATACCG	CGCAAGAAAGTCTAC	CGCCCGGACATCTATCT	CGCGCGCGCGCAAGGAACGTGTC	1218		
Db	117264	GACATTCG	CAACAGAGGTGTATGAC	ACAGATTTTATCAAC	CGCGCAAGAGCACTAATTC	1172059		
QY	1219	GCAGAGGCA	AGCGCCAGGCGCGA	GAATCTTCCCTGCC-----	GATACCTTCATCAAG	1269		
Db	117204	GAAAGAGG	AGCATATGAGCGCGCATGTATTTT	CTGACTTTGCC	AAAGAAAGCGGTTTTCCG	1171454		
QY	1270	CCGTCGCA	GAATTTCTTCAATG	CAAAAGTGC	CGTTTCATGTC	CAACAAGCCCAACGATTAAC	1329	
Db	117144	CCACCGCA	AAACCACTTCAATG	ATGATGATTC	CCCTATG	ACGCGCGTATGCCCAATATGCTAT	1170859	
QY	1330	CTCGCAAG	TTTGATGATG	GTCTGAAGGC	1359			
Db	117084	CTGACGAC	GTTTTGCTATTTG	CGCTGTGAAGGC	117055			

	RESULT 3	
LOCUS AP005348/c		
DEFINITION	AP005348	24885 bp DNA
ACCESSION	Vibrio vulnificus YJ016 DNA,	linear BCT 06-JUL-2004
VERSION	section 5/8.	
KEYWORDS	AP005348 BA000038	
SOURCE	AP005348.1 GI:37201116	
ORGANISM	Vibrio vulnificus YJ016	
	Vibrio vulnificus YJ016	

REFERENCE	AUTHORS	TITLE	JOURNAL	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL
1	Chen, C.Y., Wu, K.M., Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L., Liu, Y.M., Chen, H.J., Shen, A.B., Li, J.C., Su, T.L., Shao, C.P., Lee, C.T., Hor, L.I. and Tsai, S.F.	Comparative genome analysis of <i>Vibrio vulnificus</i> , a marine pathogen	Genome Res. 13 (12), 2577-2587 (2003)	14656965	2	(bases 1 to 248850)	Chen, C.Y., Wu, K.M. and Tsai, S.F.	Direct Submission
								Submitted (29-May-2002) Shih Feng Tsai, National Health Research Institutes, Division of Molecular and Genomic Medicine; 128, Yen-Chiu-Yuan Road, Sec 2, Taipei 115, Taiwan (E-mail: pccai@nhi.org.tw, Tel: 886-2-8146-1041, Fax: 886-2-2789-0484)
								This sequence was determined by the Sequencing Core of the National Yang-Ming University Genome Research Center (YMG), http://genome.ym.edu.tw/
FEATURES	Location/Qualifiers							
source	1. 248850							
	/organism="Vibrio vulnificus YJ016"							
	/mol_type="genomic DNA"							
	/strain="YJ016"							
	/db_xref="taxon:196600"							
	/chromosome="I1"							
	383. .1786							
gene	/gene="VVA0914"							
	383. .1786							
CDS	/gene="VVA0914"							
	383. .1786							
	/codon_start=1							
	/transl_table=11							
	/product="anaerobic C4-dicarboxylate transporter"							
	/protein_id="BAC96940.1"							
	/db_xref="GI:37201117"							
	/translation="MNRVYFPCITRNEITNITSIRINNMVLYEPLFILMTYGSRYGGIGLVSGTIGLVIEFYFKMPTSPPTVMILITAVTCASILEAGGCKITMQLQVAAERMLKPKRVTLIAPFVYTWTFMGTGSAVSIIMETIDVALKNGIRBERPAAASVAAOIIATSPISIAAVVYLAQSDINHEITLILSLVTPATLFGTLISLMSKRGKELDEDPYARLADPVMREKILNTTASIDELHPAARSVLLFISITLIVIAIIPVIRIIGETKPIISMAVVIQMMALCRGGIILATKTPRSPVGVKSGVAAIAIFGIAMSDYRQVAMPOKSGSIIVMNMYNPTPLALPIVSVNSQAQAAVARMVPLGLIEPLVLVGLMPALRGYFIIPNYSDIATVNFDTGTTKIGKTFNHSFMSVGLIGVICACCGYLLGQMIIIG"							
	1858. .1891							
gene	/gene="rplC2708"							
	1858. .1891							
repeat_region	/note="2 units of 17 bp in-tandem repeat"							
	/rpt_type=tandem							
	/rpt_unit="ttctattgtcgtcaac"							
gene	/complement (1905. .3185)							
	/gene="VVA0915"							
	complement (1905. .3185)							
CDS	/gene="VVA0915"							
	complement (1905. .3185)							
	/note="identified by GeneMark and Glimmer2"							
	/codon_start=1							
	/transl_table=11							
	/product="conserved hypothetical protein"							
	/protein_id="BAC96941.1"							
	/db_xref="GI:37201118"							
	/translation="MMKRIITLLLTQWLMQPVLAAGTNSQYARVNQNAQQLADNNLAALAKLANSFVSRPDDVALSKLIGIYYWAEQFTRISLEKALSFEITFEBEAVTFRMLTQVLTQGDYKKAIPQNKLLQVPEMSAEIWRVLAQAQSVENRRETATVATGQFQVREKKPPTTLISLAAHVQLEQKSVPLAKOILLESKNNMLQAYNATLNGQSKALDMTLTILQGLBLNATKSLATLYAAQGIYQRAQYTLAIENSOQDPRLKIQQAQVQAKEWQALETMQARLQKSSQYQWEVALLENDQRYQVYLISAEMQNPGRQDPAQLKINALYRERIDQAVQAQKADALRSTKOTQWMLKSHLKSSESTLDAKGEQVSEERHLTQQRNVVADKSDSS"							
	complement (3182. .3820)							
gene	/gene="VVA0916"							
	complement (3182. .3820)							
CDS	/gene="VVA0916"							
	complement (3182. .3820)							

gene
CDS
/codon_start=1
/transl_table=1
/product="periplasmic protein TonB2"
/protein_id="BAC96942.1"
/db_xref="GI:37201119"
/translation="MNRVIAFPPALALVFAALFSAALLVNSGKALQIDOTPSIFNNV
VMEODTEFRORSVPPEPPEIPPEOPOTPTQMTSTQVATKPIDMRFNVSGL
LAVSIATPATQNASNTTKQMLPSREAVYPAKAKRRLEGVTLKPIDERGR
KMLEVADPAPFENNADAKIKRRYQPIVDGQQTGYTTKIEFKMAQ"
/complement(3820..4224)
/gene="VVA0917"
/complement(3820..4224)
/gene="VVA0917"
/codon_start=1
/transl_table=1
/product="biopolymer transport protein ExbD2"
/protein_id="BAC96943.1"
/db_xref="GI:37201120"
/translation="MELGRQSKDEBAQIDLTSMLDIYFMLIFITSSVRESGVY
VNRQANVSAGKDAGIFVAITANDIYIDKMDVERVQASIERLLLEQDASLVIQ
ABEYANGVNVVPMAPAKAGVKNALAAEKS"
/complement(4221..4787)
/gene="VVA0918"
/complement(4221..4787)
/gene="VVA0918"
/codon_start=1
/transl_table=1
/product="biopolymer transport protein ExbB2"
/protein_id="BAC96944.1"
/db_xref="GI:37201121"
/translation="MINDHSDIPELLTSSMTNALYSTPQCGEVLVWLAAVIT
FWLTIKELFTYCAVPRIRATTLQKQWRDHLSWFRSQRKMLIDAHQILFKHLA
FKIVAVLIPMLGLGTVMGLSVDFVATQSSNPKLMSGISLATLPTAGAAVAL
VCMFVARLVKLCQREVLLEBSLRSDS"
/complement(4784..6151)
/gene="VVA0919"
/complement(4784..6151)
/gene="VVA0919"
/codon_start=1
/transl_table=1
/product="biopolymer transport protein ExbB-related
protein"
/protein_id="BAC96945.1"
/db_xref="GI:37201122"
/translation="MNKRLISLALVSLSPFAQTQVPRKATQESRQQAENQRE
NAFLSEQQLAEKROLEKQSTTEKVALISQPSDNEVTLSHLEELRRTGSLGE
LVSGRHAHKEKADLANAVTAIGOTBEVYVAVVAAQSLPDAQLKTLHWNQOQI
VASGERALQVNLINABGAVEPTDMALGNRFLALPGCFVWMSDKGALPYALPBG
AATLTLDSQHAISWPAVIDPRTGVILQYAHSPSLADRPAGGVGKALPYALPBG
IVLAVYRGILTLTTSKTRILQLEKPNQGNPLDGVILAVYDKDPRSEVLELLELT
IYDEQTEGKESIMKLIALAALPMGLGTVMGLTETFOVITQNGDPKVMASGIS
MALTIVLGVAMAPLILAHNVLSGWAERTIRSLIEKQIGLVAQAQEQQLPSLPTMDK
VGNAA"

Query Match	Best Local Similarity	Score	DB 1;	Length
Matches 867; Conservative	64.2%;	Pred. No. 5,7e-94;		
	0;	Mismatches 468;	Indels 15;	Gaps 3;
25	TCGAAAGAAATGTTATGACGTTATCGGCTTCGCTACCGCTTGGGGTTGACGATTCGC	84		
Db	TCGATGAAAAAGAGATGACGATGGCGGTATTAACAAGTTGGTGGCTTCCACCG	4680		
Qy	CCCGATGTCGGGGGAGTTGGCAAGCTGGAAGGAATTTGAAATTCGGCTTCATCAAG	144		
Db	GTTGTACTCGGAGAGTCGGCGGAGGCTGAAAAAGAGACTTAAGTTGGCTTATCAAA	4674		
Qy	CTCAACGACATGGGCGCGCTGGCGGTGGCGCGCCGCAAAAAGGCTTCTTCAAGACGAGGC	204		
Db	TTGACCGACATGGCAGCGTTGGCGCATCGCTTATGAAAAGGATTTTGTGAGATGAAGGG	4668		
Qy	CTGTTCTGTCACATCGGAAGCGCAGGCGCAACTGGAAGGTGTGATGATAGGCTTCGTAAT	264		
Db	CTGTATGTGACGCTAGGAAGCGCAGGCGCAACTGGAAGGTGTGCTCGATCGCTGATTTGAT	4662		
Qy	GGCGAATCGGACGGCTGCGACATCTGTGGCGCGCGCGCTTATGCGCGCAGCGTTGGCTTC	324		
Db	GGTAACTGAGTGTGTGGCGACATGCTGCGACGCGCAACTGAGGTGTGCACATCGGCTATT	4656		
Qy	GGCAACCAAGCCGATATCGAGGTGCGCTTCAAGCATGGGCTTCAACGCGCAACGCGATTACG	384		

46565	GGCAGCGAACCCAAAGTGATCACCGCATTCAGATGGAATCTCAATGTAACGCATACACC	46506
QY	385 GTGTCCAAATGAATCTGGCATCGATGGAAGCCCAATACCGCTGGAAG---CGGTAA	441
Db	46505 GTGTCCAAACGATGTTTGGCAGCAAAATGAAGCCCAATTTGGCCAAACAGAGTAA	46446
QY	442 CCGGTGATCCGATCAAGGAGATTAATCTAAACCGCGTGGTGAATAATATCAAGGCCGA	501
Db	46445 CCGGTTCAACCCCATCAAGGCGGATGCCCTTAAAGCTGTGTAAACGATTAATGCGACCA	46386
QY	502 GGCAGCGGTTCAATATGCGATGACCTTCCGGCCGGATCTTCAACATCAAACTGGCGT	561
Db	46385 GGCAGAACGTTCAATATGCGATGATGTTCCGGTTCAACGATTAATCAAGACTGCGCG	46326
QY	562 TACTGCTGGCGGCTGGCGGTATCAATCTGGGTACTATTTGGCGCGCA---GACATT	618
Db	46325 TACTGCTGGCGGCTGGGGGAATTCATCTGGGTATTAATGCGCGGATTAAGGTGAAC	46266
QY	619 TCCGCGCAATTCGCGCAGACGCGCTTGTGTCCGTGACCCCGCGCGCAAAATGCCGTCC	678
Db	46265 AGTGGCCAAATTATATCCGAGGTGTGTCTCAGGTGACGCTCTCCACCAAAATGCCAGCC	46206
QY	679 ACGCTGGAAGCGCGACCAATTTTCGTTATTGCGTGGCGAGCGGTGAACCAAGCGCC	738
Db	46205 ACCATGGAACGGGGTACATCAAAAGGTATTTGGTGTGATCTTGGAAACCAACAGCG	46146
QY	739 GTGTTCAAGGCGATAGCGGTGCGCGGTATCACCGATGAAGAACTTGAAGGACACGGCG	798
Db	46145 GTGTTTAAGGATTAATGCGGTGCGCGGTATCACCGATGAAGAACTTGAAGGAAACCA	46086
QY	799 GAAAAAGTCTTCGCGGTGACCAACCAATGGCGGAGAAATATCCCAACCTATCTGGCG	858
Db	46085 GAGAAAGTATTTGGTGTGTCACAAATTTGGCGGAAAGTACCCCAATACCATATTCGCG	46026
QY	859 GTGACCAAGCGCTGATTTGGGCGCGGATCTGGCTGAGCGCGCAATTAACAGAACCGCC	918
Db	46025 GTGTGTCAAAACCCCTGATCCGTGGCGGCATTTGTTAGATGAACATTAACGCAATGCG	45966
QY	919 AAGGAACCATCGAAATGCTGGCGCAAAAACATTAAGTGGGTCTGACGTGGAAGTGTG	978
Db	45965 ACCGAAGCCGTGAAGCTACTTTACGCGAGTGAATGCTGGCGCGCATGCAAGATGATC	45906
QY	979 GCGGTAGCATGAACGGCACTTTTGAATCGAAAAAGACGATTAACGCGCGCTACCGGAC	1038
Db	45905 GCCAATTCATATGACTGGCAAGTTGAATACGAAAAAGCGATTAAGCGACAAAGTCCGAT	45846
QY	1039 TTCAACACTTTTTCGCGCAACGGCGCAAGTATTCGTCTCAACGAGTGAAGTCTGGTAT	1098
Db	45845 TTTAATGTGTTCTTTGCTACAAAGCGCACTATCTCTACTACGAGCAACGGATTTGGTAT	45786
QY	1099 CTGACACAGCTGAAGCGCTGGGGCATGATCAATGAATTCAAACGGGACAACTGGTATCTG	1158
Db	45785 CTCAACAAATGCGCGCTTGGGACAAATGGAACCAAGAAACCGACTCATGTGTACATG	45726
QY	1159 GATACCGCCAAAGACGTCTACCGCGCGGACATCTATCTCGCGCGGCCCAAGAACTGGTC	1218
Db	45725 GACATCCCCAAAGCAGTGTATTCACACGATATTATCAACGCGCGCAGAAAGCGCTAATC	45666
QY	1219 GCAGAAAGCAAGCGCAAGCGCGAAGATTCCTCGGC-----GATACCTGATCAG	1269
Db	45665 GAGAAAGGACATTTGAACGCGCAAGGATTTTCCTGACTTTGCCAAAGAAAGCGGTTTTCGC	45606
QY	1270 CCGTGCAGAAATTTCTTCATCGACAAAGTCCGTTGATGCCAAACAGCCACGATTAAC	1329
Db	45605 CCAACGCAAAACCACTTCATTTGATCAGATCACTCATATGACGGCGGTAGCCCAATGGTAT	45546
QY	1330 CTGCGCAAGTTTGGATAGGTGTGAAGGC 1359	
Db	45545 CTGCAGAGTTTCGCAATTTGGCTGTGAAGGC 45516	

LOCUS	302100 bp	DNA	linear	BCT 07-APR-2004
DEFINITION	Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 4/6.			
ACCESSION	AP005087 BA000032			
VERSION	AP005087.1 GI:28809161			
KEYWORDS				
SOURCE	Vibrio parahaemolyticus			
ORGANISM	Vibrio parahaemolyticus			
REFERENCE	Bacteria: Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.			
AUTHORS	1. Nasu, H., Iida, T., Sugahara, T., Yamaiichi, Y., Park, K.S., Yokoyama, K., ..			
TITLE	Makino, K., Shinagawa, H. and Honda, T. A filamentous phage associated with recent pandemic Vibrio			
JOURNAL	Parahaemolyticus O3:K6 strains			
MEDLINE	J. Clin. Microbiol. 38 (6), 2156-2161 (2000)			
PUBMED	20295086			
REFERENCE	10834969			
AUTHORS	2. Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,			
TITLE	Tagomori, K., Iijima, Y., Nejima, M., Nakano, M., Yamashita, A.,			
JOURNAL	Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H.,			
MEDLINE	Hattori, M. and Iida, T.			
PUBMED	Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism			
REFERENCE	distinct from that of V cholerae			
AUTHORS	Lancet 361 (9359), 743-749 (2003)			
TITLE	3 (bases 1 to 302100)			
JOURNAL	Hattori, M., Yamashita, A., Oshima, K. and Shiba, T.			
MEDLINE	Direct Submission			
PUBMED	Submitted (09-APR-2002) Masahira Hattori, Kitasato Institute for			
REFERENCE	Life Sciences, Kitasato University, Kitasato 1-15-1, Sagamihara,			
AUTHORS	Kanagawa 228-8555, Japan			
TITLE	(E-mail:hattori@genome.its.kitasato-u.ac.jp,			
JOURNAL	URL:http://genome.its.kitasato-u.ac.jp/, Tel:81-42-778-8194,			
MEDLINE	Fax:81-42-778-8193)			
PUBMED	This work was done in collaboration with Kozo Makino, Ken Kurokawa,			
REFERENCE	Katsunshi Yokoyama, Takayuki Uda, Kenichi Tagomori, Masatomo Nejima,			
AUTHORS	Masayuki Nakano, Yoshino Kubota, Shigenobu Kimura, Teruo Yasunaga,			
TITLE	Takeishi Honda, Hideo Shinagawa, Tetsuya Iida (Osaka University),			
JOURNAL	Yoshio Iijima (Kobe Institute of Health), and supported by the			
MEDLINE	Research for the Future Program of the Japan Society for the			
PUBMED	Promotion of Science. This clone was isolated from a patient			
REFERENCE	presenting with acute gastroenteritis.			
AUTHORS	Location/Qualifiers			
TITLE	1. 302100			
JOURNAL	/organism="Vibrio parahaemolyticus"			
MEDLINE	/mol_type="genomic DNA"			
PUBMED	/strain="O3:K6"			
REFERENCE	/sub_strain="RIMD 2210633"			
AUTHORS	/db_xref="taxon:670"			
TITLE	/chromosome="2"			
JOURNAL	53. .1139			
MEDLINE	/note="disrupted by frameshift, similar to part of			
PUBMED	hypothetical protein GB:AAF96452.1 (AE004386) percent			
REFERENCE	identity 62 in 165 aa"			
AUTHORS	complement (1224. .2012)			
TITLE	/gene="VPA0861"			
JOURNAL	complement (1224. .2012)			
MEDLINE	/gene="VPA0861"			
PUBMED	/note="similar to GB:AAF96483.1 (AE004389) percent			
REFERENCE	identity 73 in 261 aa"			
AUTHORS	/codon_start=1			
TITLE	/evidence=not experimental			
JOURNAL	/trans_table=1			
MEDLINE	/product="conserved hypothetical protein"			
PUBMED	/protein_id="BAC62204.1"			
REFERENCE	/db_xref="GI:28809162"			
AUTHORS	/translation="MKYWKASALITLTLAGTLAGSAPPTGRNOLIMPSDOEKSISGAKS			
TITLE	FDMDKIPISDKTKTNACVCAKOTLNLIPDAGKDMEVVFDSDVNAFLPGG			
JOURNAL	KIGYITGLLVAKKQDOLATVIGHVAHLVAHSNELSSQSLANAGLSLANVAIGKS			
MEDLINE	EAYGQDLTMAAGVGVQYGVLLPYGTQSEADIVGLRYTAKAGFNPNSVDLMKMM			
PUBMED	SASGAGQPEPFSTHPSHSTRIDQATLNKLPKYNVAKPKG5"			

gene	2217. .2651	/gene="VPA0862"	/note="similar to GB:AAF96482.1 (AE004389) percent identity 75 in 144 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62205.1"	/db_xref="GI:28809163"	/translation="MSIKKKPISLILNATSRNTMMEHLQIEYTNFTDDSLSATMPCSTHROPGLHGCAGSAIVLAETIGSILANPFCSEBGYICGLILNHNHRAWMSGYIGTAPRHLGVSITQWQIEITDERDLVCTSRILTAVKQAKLA"	2694. .2927	/gene="VPA0863"	2694. .2927	/gene="VPA0863"	/note="similar to GB:AAF96481.1 (AE004389) percent identity 56 in 76 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="hypothetical protein"	/protein_id="BAC62206.1"	/db_xref="GI:28809164"	/translation="MVIERSLGIATQREIVIKLTGEMWTLQAQTDALQLLGRGANVLAHGAECCKSKLDNEEQILQLAQEIGIDIQ"	3007. .4713	/gene="VPA0864"	3007. .4713	/gene="VPA0864"	/note="similar to GB:AAF96480.1 (AE004388) percent identity 66 in 565 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="putative ABC transporter periplasmic solute binding protein"	/protein_id="BAC62207.1"	/db_xref="GI:28809165"	/translation="MSSPRLRQVETLLEKRAQNDTVEQLDITDALFCRRNARIYVNLSEEGWLEHMPAAGRGKLSLIPKRRSDVSENLARYLDEKIQGLADALDNDAA RLTVQIOGLIQLHROGQGVNRLPYRPLAMNPOKPKRSEOHIAQVFGLTRLDEBSOLQPDLAHTEALSDTHMRPYLRGVRFLNNGEPLTDCVSESILASGLNLSHQLQVSSQPPWQDILKVRPRLYPLALSESQAVTLPOALRSDDPRKPIGTSPQVKVNNDKRILLTPAGVYRGFRPLGYVFWVLDAYSSMYSLSKPPMKNQASDPEYELDPGCTPLLNKNTGIADKPRMAEFLSOTLNSHQITYAHVPQDKYTEIGVLDQAFLEKGMIDLKPTVSVEGVPQDGKSIIVAYQKQRMFPIVAKALKHLKPSVVEFLRYDTQPPNEDEVDVWAKMGILNTRNDALAGWLLDYSPIEKSAGYDFSEWATLVDQRAIHADFPAPALRGLQVLVYSCVILPMFCHWLGKVNKHGALQNAKCNALGMWFDPSNVVAKPEIDNQAE"	gene
CDS	complement (5251. .5847)	/gene="VPA0865"	complement (5251. .5847)	/gene="VPA0865"	/note="similar to GB:AAF96687.1 (AE004407) percent identity 71 in 192 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62208.1"	/db_xref="GI:28809166"	/translation="WKSSTLGLGTLIPAIYVFLSALPESSRAYWPAEIIIPATGILIAIMAIIRQFQSTAYTALNFIWLCFTLAKRTFAEVPDWMNNLIGSERNNDFVAHFAIGLVAIPAEALTIRNKKENPTFSCFPAITPAIMSLAAGYEIIEWWYELAGDGGIAPALGSGDIDWAQDMDCITTGAILSLFLMSAQRRFAKPP"	6123. .6962	/gene="VPA0866"	6123. .6962	/gene="VPA0866"	/note="similar to GB:AAF96482.1 (AE004389) percent identity 75 in 144 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62209.1"	/db_xref="GI:28809167"	/translation="MSSPRLRQVETLLEKRAQNDTVEQLDITDALFCRRNARIYVNLSEEGWLEHMPAAGRGKLSLIPKRRSDVSENLARYLDEKIQGLADALDNDAA RLTVQIOGLIQLHROGQGVNRLPYRPLAMNPOKPKRSEOHIAQVFGLTRLDEBSOLQPDLAHTEALSDTHMRPYLRGVRFLNNGEPLTDCVSESILASGLNLSHQLQVSSQPPWQDILKVRPRLYPLALSESQAVTLPOALRSDDPRKPIGTSPQVKVNNDKRILLTPAGVYRGFRPLGYVFWVLDAYSSMYSLSKPPMKNQASDPEYELDPGCTPLLNKNTGIADKPRMAEFLSOTLNSHQITYAHVPQDKYTEIGVLDQAFLEKGMIDLKPTVSVEGVPQDGKSIIVAYQKQRMFPIVAKALKHLKPSVVEFLRYDTQPPNEDEVDVWAKMGILNTRNDALAGWLLDYSPIEKSAGYDFSEWATLVDQRAIHADFPAPALRGLQVLVYSCVILPMFCHWLGKVNKHGALQNAKCNALGMWFDPSNVVAKPEIDNQAE"	gene										
CDS	complement (5251. .5847)	/gene="VPA0865"	complement (5251. .5847)	/gene="VPA0865"	/note="similar to GB:AAF96687.1 (AE004407) percent identity 71 in 192 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62208.1"	/db_xref="GI:28809166"	/translation="WKSSTLGLGTLIPAIYVFLSALPESSRAYWPAEIIIPATGILIAIMAIIRQFQSTAYTALNFIWLCFTLAKRTFAEVPDWMNNLIGSERNNDFVAHFAIGLVAIPAEALTIRNKKENPTFSCFPAITPAIMSLAAGYEIIEWWYELAGDGGIAPALGSGDIDWAQDMDCITTGAILSLFLMSAQRRFAKPP"	6123. .6962	/gene="VPA0866"	6123. .6962	/gene="VPA0866"	/note="similar to GB:AAF96482.1 (AE004389) percent identity 75 in 144 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62209.1"	/db_xref="GI:28809167"	/translation="MSSPRLRQVETLLEKRAQNDTVEQLDITDALFCRRNARIYVNLSEEGWLEHMPAAGRGKLSLIPKRRSDVSENLARYLDEKIQGLADALDNDAA RLTVQIOGLIQLHROGQGVNRLPYRPLAMNPOKPKRSEOHIAQVFGLTRLDEBSOLQPDLAHTEALSDTHMRPYLRGVRFLNNGEPLTDCVSESILASGLNLSHQLQVSSQPPWQDILKVRPRLYPLALSESQAVTLPOALRSDDPRKPIGTSPQVKVNNDKRILLTPAGVYRGFRPLGYVFWVLDAYSSMYSLSKPPMKNQASDPEYELDPGCTPLLNKNTGIADKPRMAEFLSOTLNSHQITYAHVPQDKYTEIGVLDQAFLEKGMIDLKPTVSVEGVPQDGKSIIVAYQKQRMFPIVAKALKHLKPSVVEFLRYDTQPPNEDEVDVWAKMGILNTRNDALAGWLLDYSPIEKSAGYDFSEWATLVDQRAIHADFPAPALRGLQVLVYSCVILPMFCHWLGKVNKHGALQNAKCNALGMWFDPSNVVAKPEIDNQAE"	gene										
CDS	complement (5251. .5847)	/gene="VPA0865"	complement (5251. .5847)	/gene="VPA0865"	/note="similar to GB:AAF96687.1 (AE004407) percent identity 71 in 192 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62208.1"	/db_xref="GI:28809166"	/translation="WKSSTLGLGTLIPAIYVFLSALPESSRAYWPAEIIIPATGILIAIMAIIRQFQSTAYTALNFIWLCFTLAKRTFAEVPDWMNNLIGSERNNDFVAHFAIGLVAIPAEALTIRNKKENPTFSCFPAITPAIMSLAAGYEIIEWWYELAGDGGIAPALGSGDIDWAQDMDCITTGAILSLFLMSAQRRFAKPP"	6123. .6962	/gene="VPA0866"	6123. .6962	/gene="VPA0866"	/note="similar to GB:AAF96482.1 (AE004389) percent identity 75 in 144 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62209.1"	/db_xref="GI:28809167"	/translation="MSSPRLRQVETLLEKRAQNDTVEQLDITDALFCRRNARIYVNLSEEGWLEHMPAAGRGKLSLIPKRRSDVSENLARYLDEKIQGLADALDNDAA RLTVQIOGLIQLHROGQGVNRLPYRPLAMNPOKPKRSEOHIAQVFGLTRLDEBSOLQPDLAHTEALSDTHMRPYLRGVRFLNNGEPLTDCVSESILASGLNLSHQLQVSSQPPWQDILKVRPRLYPLALSESQAVTLPOALRSDDPRKPIGTSPQVKVNNDKRILLTPAGVYRGFRPLGYVFWVLDAYSSMYSLSKPPMKNQASDPEYELDPGCTPLLNKNTGIADKPRMAEFLSOTLNSHQITYAHVPQDKYTEIGVLDQAFLEKGMIDLKPTVSVEGVPQDGKSIIVAYQKQRMFPIVAKALKHLKPSVVEFLRYDTQPPNEDEVDVWAKMGILNTRNDALAGWLLDYSPIEKSAGYDFSEWATLVDQRAIHADFPAPALRGLQVLVYSCVILPMFCHWLGKVNKHGALQNAKCNALGMWFDPSNVVAKPEIDNQAE"	gene										
CDS	complement (5251. .5847)	/gene="VPA0865"	complement (5251. .5847)	/gene="VPA0865"	/note="similar to GB:AAF96687.1 (AE004407) percent identity 71 in 192 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62208.1"	/db_xref="GI:28809166"	/translation="WKSSTLGLGTLIPAIYVFLSALPESSRAYWPAEIIIPATGILIAIMAIIRQFQSTAYTALNFIWLCFTLAKRTFAEVPDWMNNLIGSERNNDFVAHFAIGLVAIPAEALTIRNKKENPTFSCFPAITPAIMSLAAGYEIIEWWYELAGDGGIAPALGSGDIDWAQDMDCITTGAILSLFLMSAQRRFAKPP"	6123. .6962	/gene="VPA0866"	6123. .6962	/gene="VPA0866"	/note="similar to GB:AAF96482.1 (AE004389) percent identity 75 in 144 aa"	/codon_start=1	/evidence=not experimental	/transl_table=11	/product="conserved hypothetical protein"	/protein_id="BAC62209.1"	/db_xref="GI:28809167"	/translation="MSSPRLRQVETLLEKRAQNDTVEQLDITDALFCRRNARIYVNLSEEGWLEHMPAAGRGKLSLIPKRRSDVSENLARYLDEKIQGLADALDNDAA RLTVQIOGLIQLHROGQGVNRLPYRPLAMNPOKPKRSEOHIAQVFGLTRLDEBSOLQPDLAHTEALSDTHMRPYLRGVRFLNNGEPLTDCVSESILASGLNLSHQLQVSSQPPWQDILKVRPRLYPLALSESQAVTLPOALRSDDPRKPIGTSPQVKVNNDKRILLTPAGVYRGFRPLGYVFWVLDAYSSMYSLSKPPMKNQASDPEYELDPGCTPLLNKNTGIADKPRMAEFLSOTLNSHQITYAHVPQDKYTEIGVLDQAF											

Query Match	Similarity	Score	DB	Length
Best Local	64.4%	Pred. No. 3.6e-89		
Matches	827	Conservative 0	Mismatches 443	Indels 15; Gaps 3;
gene				
Db	101	TTGGCAAGCTGAAAAGAGATTGAAATTTGGCTTCATCAACCTACCGCAATGGCGC	160	
Db	217725	TGGGTGAACCGAAATTTGAAGACTGTGAAGCTTGTATTAAGTTAACGATATGGCGC	2176	
Qy	161	CGCTGGGCGGTGGCCCGCGGAAAAGGCTCTTCGAGAGCGAGGGCGTGTGCTGCAATGG	220	
Db	217665	CGTTAGCGGTGTGCTTACGAGAAAGGCTTTTTTGAAGCGAAAGGCTATTAAGCTTACGCTTCG	2176	
Qy	221	AAGCGCAGGCGCAACTGGAAGGTGTGTGATGATAGAGTCTGTAATGGCGAACTGACCGCT	280	
Db	217605	AAGCGCAGGCGCAACTGGAAGGTGTGTGATGATAGAGTCTGTAATGGCGAACTGACCGCT	2175	
Qy	281	CGACATGCTGGCGCGCGCGCTTTAGCGCGCAGCGTGGCTTGGCAACCAAGCGCGATA	340	

Db 217545 CGCATATGCTGGACGAGGCGCATTTGGGCGCAACGATTTGTCGGACCAAAAACGAG 217486
 Qy 341 TCGAGGTGCGCTTCAAGCATGGGCTTCAACGCGCAACGGATTTACGGTGTCCAAATGAATCT 400
 Db 217485 TGAATACCGGCTTCAAGTATGGAATTTGAACGCAACCGAATCACTGTTTCCAAATGACAGT 217426
 Qy 401 GGCATCATGTAAGCCGCAACATAC--GCTGGAAGCGGTAAACCGGTGATCCGATCA 457
 Db 217425 GGGAAACAGTAAGAACCAACATATCAAAAGACCGCATGTGCAAACTGTATCATCATCA 217366
 Qy 458 AGGCAATATCTAAACCGGTGTGCAAAATATCAAAAGCGAAGCGGAGCGGTTCAATA 517
 Db 217365 AAGCGATTTGTTGAAGCCCGGTGTGTATGCGGACAGAGGCAAGTCAATTCATTA 217306
 Qy 518 TGGCGATGACCTTCCGCGCGGATCTTCAACATCAACTGCGTTACTGTGCGCGGCTG 577
 Db 217305 TGGGCAATGGTGTCTGTTTCACTCAATTAAGATTGCGTTATTTGGCTTGTGCTG 217246
 Qy 578 GCGGTATCATCTCGGTACTATTCGCG--GCGGACAGCATTTCCGCGCAATCGCGG 634
 Db 217245 GTGGGATTCATCCAGGTTTATATGACACAGACTCAGCGCATTAACAGTGTCAATTGGATG 217186
 Qy 635 CAGACGCTTGTGTGGGTGACCCCGCGCGGCAAAATGCGGTCCAGCTGGAAGCGGCA 694
 Db 217185 CTGACGTTCTTTGAGTTTACCGCACCGCACCAAAATGCAAGCCACATGGAAGCGGCA 217126
 Qy 695 CCATTTTGGTTATTTGCGTGGCGGAGCGGTGAAACAGACGCGGTGTTCAAGGCGCATG 754
 Db 217125 CATTCAAAGGTTACTGGTGTGGGCAACATGGAACCGAAGCGGTATTTCAAGGCGATTG 217066
 Qy 755 GCGTCCCGTATCACCGATGAAGAACTGTGAAGAGACAGCGCGGAAAAAGTCTTCGCGG 814
 Db 217065 GTACGCGACAGTCGACCGATTAAGAAATCTGAAAGAACCAAGAAAGTGTGTG 217006
 Qy 815 TGACCAAAACAATGGCGGGAATAATACCCCAACCTATTCGGGTGACCAAGCGGTGA 874
 Db 217005 TGTCCAAAGATGGGCAAGAAAAAACCACCAACGCGCATTCGTGTGTAAGGCGCTGA 216946
 Qy 875 TTGCGGCGCGGATCTGGCTGGAAGCGCGCAATTAACAAGACCGCAAGAACCATCGAAA 934
 Db 216945 TTGCGACCGGACATTTGGTTGATGAAAAACGCAATGCGAACCCGCAAGAACCGGTGA 216886
 Qy 935 TGTGCGCGCAAAACAATACGTGCGGTGTGACGTGAAGTGTGGCGGTAGCATGAAG 994
 Db 216885 TGTGCGCAAGACGAGTATGTGGAGCGCATGCGCATTTATGCGCAATCATGACAG 216826
 Qy 995 GCACTTTGCAATTCGAAAAAGAGATTAACGCGCGCTACCGGACTTCAACCTTTCTTC 1054
 Db 216825 GCACTTTGCAATTCGAAAAAGAGATTAACGCGCGCTACCGGACTTCAACCTTTCTTC 216766
 Qy 1055 GCCACGCGCGCAGCTATTCGCTCTACAGAGTGAAGTGTGATCTGACCCAGTGAAG 1114
 Db 216765 GTGATTAACGCAAGTACCGGTACTACGCGATGCGATTTGATCTGACCCCAATGCGCC 216706
 Qy 1115 GCTGGGCGCATGATCAATGAATTAACCGGCAACTGTGATCTGATACCGCGAAGAG 1174
 Db 216705 GCTGGGCGCAAAATCTCAAGGAGAAATCGATGATGTTGATCATGAGTGTGCGAAGAG 216646
 Qy 1175 TCTACCGCGCGGACATCTATCTGCGCGCGCGCAAGAACTGTGCGCAAGAGCGCA 1234
 Db 216645 TTTACCGTCAAGATCTTACCAACAGCACTCAAGGTTAATGAAGTGTGTGTA 216586
 Qy 1235 AGGCGCAAGACTTCC-----TGCGGATACCTGATCAAGCGGTGCGAATTTCT 1285
 Db 216585 GCAAAAAAGACTTCCAGATTTCAAGTGTGAAGTGTGTTTGTGCGCCCAAAACACT 216526
 Qy 1286 TCAATGCAAAAGTCCGTTGATGCAACCAAGCCCAAGATTAAGTGTGCGCAAGTTTGGCA 1345
 Db 216525 TTAATGCAAAAGTCTTTATGAGGTGTGCGCAAGTAAATATCTGGAAGATTTCTTA 216466
 Qy 1346 TAGGTCTGAAGCAAGCAAAAGCT 1370
 Db 216465 TCGGGCTTAAAGAAAAAGCAAAAGT 216441

RESULT 5
 BX294133
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

BX294133 311050 bp DNA linear BCT 11-JUL-2003
 Pirellula sp. strain 1 complete genome, segment 1/24.
 BX294133 BX119912
 BX294133.1 GI:32442811
 complete genome.
 Rhodopirellula baltica SH 1
 Rhodopirellula baltica SH 1
 Bacteria; Planctomycetes; Planctomycetaceae; Planctomycetales;
 Planctomycetaceae; Rhodopirellula.
 1 (bases 1 to 311050)
 Gloeckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T.,
 Ludwig, W., Gade, D., Beck, A., Borzym, K., Heilmann, K., Rabus, R.,
 Schlesner, H., Aumann, R. and Reinhardt, R.
 Complete genome sequence of the marine planctomycete Pirellula sp.
 strain 1
 Proc. Natl. Acad. Sci. U.S.A. 100 (14), 8298-8303 (2003)
 2735913
 12835416
 2 (bases 1 to 311050)
 Kube, M., Borzym, K., Heilmann, K., Klages, S., Margardt, I.,
 Lehnack, S., Beck, A., Pawlik, R., Reinhardt, R., Gloeckner, F.O.,
 Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Rabus, R.,
 Schlesner, H. and Aumann, R.
 Direct Submission
 Submitted (21-JAN-2003) Max Planck Institute for Molecular
 Genetics, ProScience Innesstrasse 73, D-14195 Berlin, Germany Max
 Planck Institute for Marine Microbiology Celstusstrasse 1, D-28359
 Bremen, Germany
 This project was carried out by
 *Max Planck Institute for Molecular Genetics, Berlin, Germany; *Max
 Planck Institute for Marine Microbiology, Bremen, Germany; in the
 framework of the RBGK-project, <http://www.regx.de> -----
 Genome Center
 Center: Max Planck Institute for Molecular Genetics
 Center code: MPIWG
 ----- Summary Statistics
 Sequencing vector: pUC19; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 7142841 bases at least Q40
 Consensus quality: 7145138 bases at least Q30
 Consensus quality: 7145484 bases at least Q20
 Quality coverage: 8.03

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid sequence; assembly was additionally confirmed by long
 range PCR and cosmid end sequences.

See <http://www.micro-genomes.mpg.de/pirellula/> for more information
 including minimal tiling path from a set of 220 cosmids out of
 908. See the misc_feature tag below for the boundaries of the MTP
 cosmids. ----- Annotation
 Center: Max Planck Institute for Marine Microbiology
 Celstusstrasse 1, D-28359 Bremen, Germany.
 Center code: MPIWG
 Email: fog@mpi-bremen.de
 Phone: +49 (0)421 2028 938 Fax: +49 (0)421 2028 560

Three different programs (Glimmer, Critica and Orpheus) were used
 for ORF-prediction. A nonredundant list of ORFs was generated by
 suitable parsing of the results.
 Automated annotation was done with the software package Pedant Pro
 (<http://www.biomed.de>). All ORF predictions and annotations were
 manually corrected by considering all results of the different
 tools applied. See <http://www.regx.de> for more information and


```

Db      184884 GAGGACGTGACGTCCTTGGAAACCCAGTCCGAACTGGAAGATCCTTTCGACAAAGTC 184943
Qy      259   GGTAAATGGCAACTGACGCGCTCGCACATGCTGCGCGCGCGCGCTTAGCGGCGAGGTT 318
Db      184944 ATCAATGGGCAACTGCGCGGCGGACACATGTTGGCCGGCCAAACGATCGGTGCAACGATC 185003
Qy      319   GGGTTCGGGACCGAAGGCGGATATCGAGGTGCGCTTCAGCATGGGCTTCAACGGGCAACGG 378
Db      185004 GGTGTGGGAACCAAGTCGCCAATCGTGACCGCTTACAGCCTTGACATCAACGGGCAATGGC 185063
Qy      379   ATTACGGTGCATGAAATCTGGCATCAGATGAAGCCGAACATAC---GCTGGAAGGC 435
Db      185064 ATCAACGGTCGACGACGAAGTTTGGGACACAGATGACGAGAAACGATCTTGAATGGAAGC 185123
Qy      436   GGTAAACCGGTGATCCGATCAAGGACGATTAATCTAAACCGGTGTCGAAAAATAC--- 492
Db      185124 CCGACACCGAAACATCCCATCAGGGCGCGACGCTGAACCGATGTGCGAGATTACCTG 185183
Qy      493   AAGCGGAAGGCAAGCCGTTCAATATGCGCATGACCTTCCCGCGGATCTCACACATC 552
Db      185184 CAAAGCGCGGCGCAACCATTCGCGATGGCATGGTGTTCCTGTACACTCAATTAAC 185243
Qy      553   AACTGGTTACTGCGGTGGCGGCGGTGATCAATCTGCTACTATTGCGCGCGGAG 612
Db      185244 GAGATTGATATTGGCTGGCGCGGTCAAGGATTCATCCGCGCATGTACACCGAATC--T 185300
Qy      613   GACATTTCCGCGCAAAATCGCGGACGACGCTTGTGTGCGTGAACCCGCGCGGCAATG 672
Db      185301 GACATCAAAAGGCTTCAACGACGACAGGTCAAAATGTGTCAGCGGTGACACCGCTCAATG 185360
Qy      673   CGGTCAACGTGGAAGCGGACCAATTTGCTTATGCTGCGGTGCGGACCGGTGGAACAG 732
Db      185361 CCTCAGAACTCGAAGGCGGACATGTCMAAGTTTACGCTGTGCGAAGCGTGAACAA 185420
Qy      733   CAGCGGTGTTCAAGGCGATAGGCGTGGCGGATGACCGATGAAGAACTCTGGAAGAC 792
Db      185421 AAGCGGTGTGACTGGCATCGGTGCTGATCCGTCAGTCCAGCACTACGATCTGGAAGAC 185480
Qy      793   ACACCGGAAAAAATCTTTCGCGCTGACCAAACTAGGCGGAGAAATACCCCAACCTAT 852
Db      185481 AACCTGAAAAAGTCTTTCGCGCTGACCGGAAATGGCGGAGAAATACCAAACTCAC 185540
Qy      853   CTGGCGGTGACCAAGGCGCTGATTCGGGCGCGGATCTGCTGGAACGCGCAATTAACAG 912
Db      185541 TTGGCGGTGATCAAGGCTGATCCGTCGCGCAAAATGCTGATGCGACCGATGACTCG 185600
Qy      913   -----AACCGCAAGAAAGCCATCGAAATGCTGGCGCAAAAACATATAGTGGT 960
Db      185601 GGCATAATGCTCAATGCGCGAAGAGCGGTGAAATCTCAAGCGCAAAAGATTATGTCGT 185660
Qy      961   GCTGACGTGGAATGCTGGCGGTAGCATGAAGGCACTTTCGAATTCGAAAAAGACAT 1020
Db      185661 GCGGAGAAAGATGATTTGGCAACTGATATGGGCAAGTTTGTCTTTCATTCACCGAC 185720
Qy      1021   AAAACGCGCTACCGGACTTCAACACCTTCTTTCGCCACGCGCGACGATATCCGCTTAC 1080
Db      185721 GTTCGGAATGCGCGACTTCAACGATTTCTTCAAAACAGAGCGAGCTACCGGCACTAC 185780
Qy      1081   AGCAGTGCAGTCTGTGATCTGACCCAGCTGAGGCGCTGGGCGCATGATCAATGAATCAA 1140
Db      185781 AGGAGTGCAGTCTGTGATCTGACCCAGCTGAGGCGCTGGGCGCAATCAAGATCCAAA 185840
Qy      1141   CCGGCAACATGATCTGATACCGCGCAAGAGTCTACCGCGCGGACATCTATCTTGCC 1200
Db      185841 CCGGCAAGTTGATACCGCGCAAGAGGCGCGCAAGATCTCAAGCTGTGATCTATCGCAA 185900
Qy      1201   GCGGCAACAGAACTGTGTCGAGAGGCAAGGCGCAAGGCGCAAGACTTTCCTGCG 1253
Db      185901 GCGGCGAGGCTTCTAATCAAGCAAGGCAACTGATCAATATGAATATCCGGC 185953

```

RESULT 6
AB017282_06/c

```

MPCOMMENT
Sequence split into 33 fragments
Fragment Name      Begin      End      LOCUS AE017282 Accession AE017282
AE017282_00      1      11000
AE017282_01      100001    210000
AE017282_02      200001    310000
AE017282_03      300001    410000
AE017282_04      400001    510000
AE017282_05      500001    610000
AE017282_06      600001    710000
AE017282_07      700001    810000
AE017282_08      800001    910000
AE017282_09      900001   1010000
AE017282_10     1000001   1110000
AE017282_11     1100001   1210000
AE017282_12     1200001   1310000
AE017282_13     1300001   1410000
AE017282_14     1400001   1510000
AE017282_15     1500001   1610000
AE017282_16     1600001   1710000
AE017282_17     1700001   1810000
AE017282_18     1800001   1910000
AE017282_19     1900001   2010000
AE017282_20     2000001   2110000
AE017282_21     2100001   2210000
AE017282_22     2200001   2310000
AE017282_23     2300001   2410000
AE017282_24     2400001   2510000
AE017282_25     2500001   2610000
AE017282_26     2600001   2710000
AE017282_27     2700001   2810000
AE017282_28     2800001   2910000
AE017282_29     2900001   3010000
AE017282_30     3000001   3110000
AE017282_31     3100001   3210000
AE017282_32     3200001   3304554
Continuation (7 of 33) of AE017282 from base 600001 (AE017282 Methylococcus capsulatus str.
Query Match      12.3%; Score 171.2; DB 1; Length 110000;
Best Local Similarity 53.3%; Pred. No. 8,4e-23;
Matches 542; Conservative 0; Mismatches 408; Indels 66; Gaps 6;
Qy      125   TGAATTTGGCTTCATCAAGCTCAGCAGCATGCGCGCTGGCGGCGGAAAAAG 184
Db      18318 TGAACGTGGTTTCATTCCTCCTCAGGACTGCGCTCGGTGTCATGCGCTGGAAGAAAG 18259
Qy      185   GCTTCTTCCAGGACGAGGCGCTGTTGTGTCACATGGAAGCGGCAATGGAAGTGG 244
Db      18258 GTTACTTCCGCGACACGCGCTGACGTCACTGTGCGCGCACGCTGCGGCCA 18199
Qy      245   TGATGATAGGAGTGTGGAATGGGAACTGGAAGGCTCGCACATGCTGGGCGCGCGCT 304
Db      18198 TCGCGACAAAGTGTGACACCGGCTGTGGAAGGCGCCACATGCTGCGCGATGCGGA 18139
Qy      305   TAGCGGCAAGCTTGGCTT---CGGCAACCAAGGCCGATATGAGTGGCTTTCAGATGG 361
Db      18138 TCGCGACACACCTTCGCGCTGGGGGGGGTACGGAAGGTACCATTTGGCGGATGCGTGG 18079
Qy      362   GCTTCAACGGCAACGGAATTAAGGTGTCGATGAAATCTGGCATCAAGTAAAGCCGAACA 421
Db      18078 ATCTGATGGGAACGCGCTGACGCTGATCCAAACCGCTGTAC-----GGCGCTCA 18028
Qy      422   TACCGTGAAGGCGGTAAACCGGTGATCCGATCCGATCAAGGAGATTAATCTAAACCGGTG 481
Db      18027 CCGCGCGGCGCGGCTGAGCTGCGCCAGAGGCGCGGTTTGGCGCGGCTGCTCGGGGTT 17968
Qy      482   TCGAAAAATACAAAGCGGAAGGCAAGCCGTTCAATATGCGATGACCTTCCGCGCGAT 541
Db      17967 TCGTGAACAGGACGCGGAGGTCTCTCGGTTGGCGACAGGATTCCTCTTCTC-----T 17914
Qy      542   CTCACAAACATCAACTGCTTACTGCTGCGGCTGGCGGTGATCAATCTCGGCTACTATT 601
Db      17913 CGCACAACTCACTGCTGCTGCTATTTGGCTGGCGCGCGGGGTGATTC----- 17865

```

QY	CGCGCGCAGAGACATTTCCGGCCAAATCGGCGCAGACGGCTTGTTGTGGGTGACCCCGC	661
QY	602	
Db	17864	-----CGACGCCACGTCGGCAGTACGTATTCCTCGC
QY	662	CGCCGCAAAATGCGTCCACGCTGAGAACCGGACCATTTTCGTTATTCGTGCGGAGC
Db	17832	CGCGCGCAGATGCCAGTTCCTGGCGGGGCGAGATTTCCGGCTACTGGCGGGGTGAAC
QY	722	CGTGGAAACGACGAGCCGTGTTCAAGGGCATAGCGCTGCCGTATACCGATGAAGAAC
Db	17772	CTGTGAAATGCCCGGTGCGGTGCCACGAGGAGATCGGCGGTGATCACCGGTATGA
QY	782	TCTGGAAGSACACGCGCGSAAAAAGTCTTCGCGCTGACCAAAACATAGGCGGAGAAATAC
Db	17712	TCTGGAAACACAGTCCCGAAAAATGCTGGGCGCTCAACACAGACTGGCGGAACCATTC
QY	842	CCAAACACTTATCTGCGGGTGAACCAAGGCGCTGATTTCCGGGCGCGATCTGGCTGAAC
Db	17652	CCAAATACCTTGGCGAGCCCTCTCTATGGCCCTGTGAGAACCGCGCAATGAGCT-----CG
QY	902	ACATTAACAGAACCCGCAAGGAAGCCATTCGAAATGCTGGGGCAAAAACATATAGTGGT
Db	17598	ATCGGCCCGAAACCGGCGCGGAACCGCGCCCTCTCCGCGAAATCCCGACTATGTGGCG
QY	962	CTGACGTGGAAGTCTGCGCGCTGACATGAACGGCATTTTCGAATACGAAAGAGAGATA
Db	17538	TGCGCGGGAGATCTCCAGATGTCATGACGGGTTCTGTTCAATATGCGGGCGGAGAC
QY	1022	AACGCG---CGTACCGGACTTCAACACTTCTTTCGCCACGGCGCCAGCTATCCGTCT
Db	17478	CGGCGGTGTCCTCCCGACTTCAACGTTGTTACCGCTATGCGGACCTTCCCTGGC
QY	1079	ACAGCAGTGCAGTGTGTATCTGACCCAGCTGAGGCGCTGGGCGCATGATGAATGA
Db	17418	GTTCCACGCGCTGTGTCTTGAACCAAGATGCTGCGCTGGGCGAGATGCCGAA
RESULT 7		
AE008293	10676 bp	DNA linear BCT 28-MAY-2004
LOCUS	AE008293	Agrobacterium tumefaciens str. C58 linear chromosome, section 97 of 187 of the complete sequence.
DEFINITION	AE008293 AE007870	
ACCESSION	AE008293.1	GI:15159385
VERSION		
KEYWORDS		
SOURCE	Agrobacterium tumefaciens str. C58	
ORGANISM	Agrobacterium tumefaciens str. C58	
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.	
AUTHORS	1 (bases 1 to 10676)	
TITLE	Hinkle,G., Slater,S.C. and Goodner,B.	
JOURNAL	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 10676)	
TITLE	Hinkle,G., Slater,S.C. and Goodner,B.	
JOURNAL	Direct Submission	
COMMENT	Submitted (14-06-2001) Bioinformatics, Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA	
FEATURES	Approximately 800 bp of telomeric sequence missing from the left end of the chromosome and 200 bp missing from the right end.	
source	Location/Qualifiers	
	1..10676	
	/organism="Agrobacterium tumefaciens str. C58"	
	/mol_type="genomic DNA"	
	/strain="C58"	
	/isolate="Cereon"	
	/db_xref="taxon:176399"	
	/complement(93..1352)	
	/gene="AGR_L_1874"	
	complement(93..1352)	

gene /gene="AGR_L_1874"
/note="HYPOTHEICAL 40.1 KDA PROTEIN IN SIPU-KIPI
INTERGENIC REGION"
/codon_start=1
/transl_table=11
/product="AGR_L_1874p"
/protein_id="AAK89510.1"
/db_xref="gi:15159386"
/translaction="MERKSMEOKKLPTPADLKMSKASLIGALFILMATSALGPFIT
OTAPFTTOLGAFAFGILASLIDPVDVQWIMRWITVLTPRNASIDANAAIPGAGYFLA
ITIVIGCGEPNIGIGTGGLINAFICIDPKTIGASISATISAI FVSRKAGIDPREIT
ITAGILMYLTIYAFVSAPPYGDAPROTFLPDTINATITITTYGTVGGYITYSGNH
RLDGGKIDLEFPVNRPAULGLAVTGMRYVFLAILGVASGVITDSGOALNVA
AQAFFSAGDIEFLRFGIIFWAAIATSVIGAYTSVSEFLPVKQDMSERANMAVAF
IATVSVCTLLITTPPAALVFGVGLNGLIPIIGISIFLFAWKREDIMGGHRSRILL
GLVTCALITWIMYKSAGTIFGLLGL"
1517. 2266
/gene="AGR_L_1876"
1517. 2266
/gene="AGR_L_1876"
/note="(ALJ33213) putative Gnr-family transcripional
regulator."
/codon_start=1
/transl_table=11
/product="AGR_L_1876p"
/protein_id="AAK89511.1"
/db_xref="gi:15159387"
/translaction="MCYGOAMPCEPMPTVEPMLHCSSERSAMTETIEGLPLADRLAK
GRTLLTIGSELAPGQRLSEAAFSPEFDSRSLSEAPRLTJKOILHEANRGVFAVW
PMSSVIDIYVRMRVCEGLAROMHHGPATKAMRAVLEAKTIRREBADMIGSAAW
KHAHAVDLODSARISDPYERILAAELHLCREBLADPQOLHSFPDMNSKILSLMBEKK
AOEAGNEMLTYNLSERTVTLKALERAEBHS"
2341. 2955
/gene="AGR_L_1879"
2341. 2955
/gene="AGR_L_1879"
/note="response regulator nast - Azotobacter vinelandii"
/codon_start=1
/transl_table=11
/product="AGR_L_1879p"
/protein_id="AAK89512.1"
/db_xref="gi:15159388"
/translaction="MLRCVINADMTLRLDTLIVIDENAIRASIIIEGLREAGORLV
IHEVNGVARTIETLOPDYITFDLENPRNDMEHLFOLTRVGRIRIAMFVDRSTASVE
AAVEGSAVYVDGKSKRVKPIIDMAVSRNAPSRLQRELDKSLKEKVIYERAK
GLIMKRGSLSEEBEAFALLROTANMEKKKISISIAQSVTAAGLLM"
2991. 4286
/gene="AGR_L_1881g1"
2991. 4286
/gene="AGR_L_1881g1"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_1881g1p"
/protein_id="AAK89513.1"
/db_xref="gi:15159389"
/translaction="MAAEKTRAGSKGATOWRDA PALVGSDDROKILPAGTIPVADYAS
LIAAEFFAPRBEGLIDLVKDVSWANVRDLARQDIAHMSPEMVASMLGSGNP
SEPTITPFLGKGNALITLSTRLFAKMAKLTGLSTPAALNARLKLVLDDMKRGEA
PPTLTMTYVPSFSHVEFRYVLAAGIHDHDKLVVPPPMTSDALAAGAIDPFCVGA
PMNIIYAARGGRIYIAAKODLMPAPKEKVIQMPREMASSOOETVIGRLITLDAAGC
DLADNHDLSGALDPRYIGAPOSIIRRVLAGERSIDSGNRRIYIEKYPTFHGDHANY
PROGSLMTYSGOMLRWGQAEISGTGUAALISAVYPDIVRALDQKRGADRIIRISQ
DEGDFVDFGPFDPDADINGVYNSFAVTSVPSAPOSEA"
4544. 5959
/gene="AGR_L_1886"
4544. 5959
/gene="AGR_L_1886"
/note="nitrate-binding protein ntrA precursor, periplasmic
- oscillatoriacean cyanobacterium"
/codon_start=1
/transl_table=11
/product="AGR_L_1886p"

```

gene
/sequence="AAK89514.1"
/db_xref="GI:15159390"
/translation="MLPSTTPVPAKQTPSGVFIFPRFRKRRHPRKMLKNGMRRK
ISGTVSRRTILKTATAALVTAVTAFPSGAPATAEPVCAKLGFIALTDAALPI
IAEKGLPAKHGMPEVEVLKQASMGATRONLVVGAASNGIDGAHILLTPMIVMHTKV
TONNVVPMPTIILARILNDSOGISVAKEYATGQDLSKAKAFKKKADGKIKYAM
TFGGHDMWIRYLAAGIDPDKDVSTIVPPVAVANKVGMVPCVGEWNNOL
VNOGIGFTACTTGBELWKGHEKALGRADVERNPATAKALVMAEQMCDPMANK
EEMSTILGKQWENVPKDVGLKLNKINGNRVLNTGLQMKFWQDHASYPFHSID
SMPTENIRMGKPAPTDVKALVAKVNRDINRAAKDLGVADLPASTSRKETPFDDG
KVDPENPSAYLESLSIKAA"
6067..6366
/gene="AGR_L_1888"
/sequence="6067..6366"
/translation="NITRATE TRANSPORT PERMEASE PROTEIN NRTB"
/codon_start=1
/codon_table=11
/product="AGR_L_1888p"
/protein_id="AAK89515.1"
/db_xref="GI:15159391"
/translation="MEMSALARKENBEPQPAVOKANVVALSPKHPRIDLRKMGAA
ARHTIPVIVLVLGVWOLCSAPGSSLPSPQVPEESVDLIVSPFHNQSDIGL
LRVLVSLERVYGFGLAINGITLCAVIGQSVAMRGDPTFOVLTVPPLAETPLSL
AAQDSNPSAIFVIFTSLMPVILINAVGRNIPDVRVAKVLRNLQLEFFKMLP
SAAPYFTGIRIGIGSLWLAIVAAEMLTGVGIGFIWDMANSSRLPDIIVALAYIGV
VGFMLDLVLAALKVITRGAAAN"
6980..7777
/gene="AGR_L_1890"
/sequence="6980..7777"
/translation="NITRATE TRANSPORT PROTEIN NASD"
/codon_start=1
/codon_table=11
/product="AGR_L_1890p"
/protein_id="AAK89516.1"
/db_xref="GI:15159392"
/translation="MNAVYLDLHIDRYFDKGSRAVYKNGINTIEKGEFVYVIGSG
CGKSTMLNLAGLTKVAGAVLLENREVEPGERAVPQNHSLPMLSYEBVNIAV
AKFRKTKRTRHDWVNRMLDVOMAHADKKSPISQMKORVGIARLAMEPIL
LIDEPFGALDALTRAHLDQAVMEIHAHLKSTMTMITHDDEAVILSDRIYMTNGRAA
HIGEVLDVPLARPRNRIELASDKIYKCAVAIKFLYERHFEAAE"
7714..10230
/gene="AGR_L_1891"
/sequence="7714..10230"
/translation="Pseudomonas aeruginosa (strain PAO1)"
/imported - Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/codon_table=11
/product="AGR_L_1891p"
/protein_id="AAK89517.1"
/db_xref="GI:15159393"
/translation="MPRGCAEVLPLRTSPLRSRGVTITOKLVIIGNGAPGMLNENL
FETAPGLYDTVFENAPRVNYDIRIMLSPVLSGEKSYEDIVIHNDVYAAANNVTLHKA
KYRIRIRANKVITSENGITPASYDLVLIATSLPFIIPBQHGOLADADLDVTKM
LEIAKKGKAVIVGAGLGLBAAYGKROGMDVTLPTIMEROLDVPAAYLLSKA
LNERGDIITKANTKCIIGBEKVEGIELENGRIIKGDMVMAVGIRIPASGLAEAGIA
VNRGIYVDGMMTSDASTYALGECABERGCYGLVAPLYBSAAYLADRLCGSABETHG
SVYNTKLVKTVGINLFSAGDPAEGDDREBYLRDATAGVYKRLILKENRIIGALVSET
ADDSWFDLMKSTDISAMRETLIFQAYGSGPLDPMVAVALPDAELCGNGVCK
GKTSATISGSLDSDVRAHTKASACGCTGLVQMLTLTGDTYNPAAYPMGCK
TDLGHDVRLRIKAKGLKTIIPAYMOLEWNTSCGACRPAALVYVCDPEPDYADY
OSQYINERVYANIOKDGTSYVPRMGVYNADELALADYNDKREPIPLVKYTGORI
DLGITEKEDIPAYMADLGRAGPISGQAYAKGLTCTGEOGKREPTQDSTGLGRI
BKPMGSWTPAKKLAIVSGCPNCAATCDDIVTCVDSGFEIHFAGALDITGRIE
LGIWTEDEALBEHIVALTOMYRQARYLEIYIMAKRIGIDVEYROIIMDADKRAYF
DRFVFOKPAQVPMSESRVSGHDHFRPMAAIGFSEAAE"
10237..10571
/gene="AGR_L_1892"
/sequence="10237..10571"
/translation="NITrite reductase small subunit PA1780"
/imported - NITrite reductase small subunit PA1780

```

ORIGIN

Query Match	9.6%	Score 133;	DB 1;	Length 10676;
Best Local Similarity	50.2%	Pred. No. 1.9e-15;		
Matches 572;	Conservative	0;	Mismatches 495;	Indels 72;
			Gaps	7;
Qy	CGATGTCGGGCGAGTTGGCAAGCTGGAAGAAAGATTGAATTGGCTTTCATCAAGCT	146		
Db	4768 CGCTTTCGGGCGACCCGCAAGACCGAAGTGAAGGTGCAAACTCGTTTATCCGCT	4827		
Qy	147 CACCGCATATGCGCCGCTGCGGTGCGCGCGAAGAAAGCTTCTTCAGAGCAGAGGCT	206		
Db	4828 CACCGACGCGCGCGCTCATCTTCGCGCGAAGAGGACTTTTCGCCAAGCATGGCAT	4887		
Qy	207 GTTTCGTCGAATGGAAGCG--CAGCCCACTGAAAGTGTGATGATGAGTATGGTCTGAA	263		
Db	4888 GCTCGAGTTCGAGGTTCTGAAACAGGCTCTTGGGGGGGCGACCGCGACATCTCGGCT	4947		
Qy	264 TGGCGAATGGAAGGCTGCGCAATGCTGGCGCGCGCGCTTACGCGGCGAGGCTTGGCTT	323		
Db	4948 CGGCGG-----CGCTCGAAGCGCATGATGGCGGCGCATATCTTAAGCCGATGCCCTTA	5001		
Qy	324 CGGCAACAAAGCGCATATCGAGGTGCGCTTACAGATGAGCTTTCAGAGCGCATATAC	383		
Db	5002 TCTCATGCAATACCGGCAAGTGAACGAACAAAGTGGCGGTCCGATGACATCTCGC	5061		
Qy	384 GGTGTCGAATGAATTTGGCATGATGAAGCCGAACATACGCTGGAAGCGGTAAAC	443		
Db	5062 ACCGCTTAATCTCGAAGAGCGAGGCGAT-----TTCGTCGCGCAAGAAATATCAGAGAC	5115		
Qy	444 GGTGATTCGATCAAGGAGATTAATTAACCGGCTGTCGAAATAACAAAGCGAAG	503		
Db	5116 CGGCTGCACTGATGATGCTCCCAAGCTGAAGGCGCTTTCGAAATAAGAGCGCGAAG	5175		
Qy	504 CAAAGCGTTCAATATATGCGATGACCTTCCGCGCGGATCTCAACATCAATCAATCGGTTA	563		
Db	5176 CAAAGAAATCAAGCGCGCATATGACTTCCCGCGCGCAAGCATGATCTGATCCGCTA	5235		
Qy	564 CTGGCTGGCGGCTGCGGTATCAATCTGAGCTACTATTGCGCGCGCAGAGCATTTCCGG	623		
Db	5236 TTGGCTGGCTGCTGCTGATCATCGATCCGACAAAG-----	5269		
Qy	624 CCAATTCGGGCGAGAGCGCTGTGTTCGGGACCCCGCGCGCAATGCGCTCAAGCT	683		
Db	5270 -----GAGCTTCACCAATCGTCTTCGCGCGCGCAATGTCGCAACAT	5316		
Qy	684 GGAAGCGGCGACCATTTTGGTTATTCGTCGCGCGAGCGGTGGAACAGAGCGCGTGT	743		
Db	5317 GAAAGTCGCAATATGAGATGCTCTTCTGTAGGCGAGCGGTGGAAGAGAGCGTGTAA	5376		
Qy	744 CAAAGGCATATGAGCGTCCGCTGATCAACCGATGAAGAACTCTGGAAGAGACGCGGAAA	803		
Db	5377 TCAGGGCATATGAGCTTATGCTGCTGATCAACCGGGGAACTCTGGAAGAGGATCCGAAA	5436		
Qy	804 AGCTTCGCGGTGACCAACAAATGGCGGGAATATCCCAACACTATTCGCGGTGAC	863		
Db	5437 GGGCTGCGATGATGCGCGGATGAGTGAAGAAAGAACCCCAAGCCCAAGGCTTTGCT	5496		
Qy	864 CAAAGCGCTGATTCGCGCGCGCATCTGCGTGAAGCGCGCAATTAACAAGAACCGCAAGA	923		
Db	5497 GATGCGCTGATGAGAGCGCGCAATGCTGCGACAGATGCGCAACAAGAGAGAGATGTC	5556		
Qy	924 AGCCATGAAATGCTGCGCGCAAAACAAATACGTGGTGTGCTGAGAAATGCTGCGCGC	983		

FEATURES	source
gene	<p> <code>/organism="Agrobacterium tumefaciens str. C58"</code> <code>/mol_type="genomic DNA"</code> <code>/strain="C58"</code> <code>/isolate="U. Washington"</code> <code>/db_xref="taxon:176299"</code> <code>complement(138..473)</code> <code>/gene="nirD"</code> <code>/note="synonym: Atu3901"</code> <code>complement(138..473)</code> </p>
CDs	

gene
/gene="ntrD"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="nitrile reductase small subunit"
/protein_id="AAU44709.1"
/db_xref="GI:17742340"
/translation="MDSNWIDIGSDIPLVNGKCVKTPGKIAVFRTAENOVAIEDHCPHAGGPELSOGIIVHGA SVTCPLHNWVISTEGALGADBEQVPTIPVRINDGLSIA
LETLIAAE"
complement(479. .2926)
/gene="ntrB"
/note="synonym: Atu3902"
complement(479. .2926)
/gene="ntrB"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="nitrile reductase large subunit"
/protein_id="AAU44710.1"
/db_xref="GI:17742341"
/translation="MTOKLVTIIGNGMARGRLLENLEFAPGLYDVTIPNAEPNVNDR
IMLSVYLGEKSEYEDIVIHNDERTYANNVTIHKAKATRIDRAKIYTSNGITASVD
RLVIVTSLPTIIPVGHQLPGVALRYRLDVTIWMLETLAKRGAIIVGALLDELEAA
YGLKQKQNDVVTIHLMPITMEROLDPAAVYLLKALNRGIDITTKANTKCIIGEEK
EGLIEDERIIKGDWVVAVGIRPAPSGAKSAGIAVNRGIIVDDGMMTSDAISALGE
CAEHGKCGIYVAPLYESARVIALDRLCGSAEYHSYNTKLTATGYNLFSAGCPAGC
DDREIVIRLDATAGVYKRLIKENRIIGAVLYGRTADGQWFFDLMKSTDISAARETLL
IRGQAVOGSGPELDPMHAAVALPDPAEITGCNCKKGTSTSIITSGKLTSLDDVAHKK
ASASGNGCTGLVPLDMLITLSDTINPAVQPKCTDIGHDVRRLTKAGLKTIPAV
RMQLEWKTSSCGACRPAALNVYLVDFEDEVADYQSKYINERVHANIQKDGYTSVAP
BMELEWKTSSCGACRPAALNVYLVDFEDEVADYQSKYINERVHANIQKDGYTSVAP
SQAQAKGLRTVTKTCVEGOMCRFGTODSTGIRLEKMMSTPAPKLKLA VSGCPAR
CAEATKCGIYVCSGSGEIHFGAGAGDITGTEVGLGVHTEDELHIALVOMYER
QARYIERIYKNAKRTGYDEVRQIMDDADRKKAIFDRFVSGKAYDPMSEVSGHD
KHEPRMAIGSEAAE"
complement(2932. .3729)
/gene="ntrC"
/note="synonym: Atu3903"
complement(2932. .3729)
/gene="ntrC"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase
protein [nitrate]"
/protein_id="AAU44711.1"
/db_xref="GI:17742342"
/translation="MNATLKLDDHIDIKYFDKGSRAEVLKGINLTIERKEEFSVIGSHG
CGKSTMLNLIGLTVTSAGVALLENREVENEGPRAVPOHSLPMLSYVENNLVA
AKVPNTPTPRERHDWWRNLDIOMAAHKQKXSEISGCKQKQVGIARLAMPKTLI
LLDEPRGLDLTPRAHLDODAYMEIHAELGNTMVTITHDVDEAVLISRIYMMNTGPRAA
HIGEVLDVLPAPRRRIELASDKIYLLKCREAVLFLYERHFEAAE"
complement(3743. .4636)
/gene="ntrB"
/note="synonym: Atu3904"
complement(3743. .4636)
/gene="ntrB"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, membrane spanning protein
[nitrate]"
/protein_id="AAU44712.1"
/db_xref="GI:17742343"
/translation="MSALARKNEEPOOPAVOKANVVALSPKHLRIDLRKMGAAAR
NTIPIVAVYVIGLIGLCAVIGOSVYANRGADPIROYLRTVTPPLMDPLSTLAA
VIVSERIYVAVFGALNIGITLIGAVIGOSVYANRGADPIROYLRTVTPPLMDPLSTLAA
FQDSPSARIPIVITSTIIPVITINAVGVRNPQDYRNVAKRLNQLEFFFKTMLPSA

APITGCRIGIGLSMLATVAEMLTGVCIGFFIWDANSSRLPIITVALAYIGVG
FMDKLVAAIGKXITRGAAAN"
/complement (4750. .6042)
/gene="ntrcA"
/note="synonym: Atu3905"
complement (4750. .6042)
/gene="ntrcA"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, substrate binding protein
[ntrcA]"
/protein_id="AAL44713.1"
/db_xref="GI:17742344"
/translation="MKKIFSGTSSRRITLKTATVAALVAVRTAPSPAPATAPEV
KAKIGPILTDAPLIIAEKGLPAKMGREVEVLKQSMGATPRLNIVIGASNGD
GAILTPRLMHTKVTQNNVPTIILRLDSDGISVAGEYATGTQULDSKIK
AAFERKADGKEIKAAATPEPGTHDLIRIWLAAAGIDPDKDVSTIVPPQVANKK
VGMADVCEGEPNBEQIVNOIGITACTGELWKGHEKALGRADWVEKNPATLAL
LMAVBAQOQCDPMANKEEMSTILGKQMFNVPPKDVGLKGNINYNGRVLENTGL
OMKFQDDHASYPHSHDSWEITENIRMGKAPPTDVKALVAKVNRD IWRAAKDG
ADLPASTSGKTEFPDGKVPDPENPSAYLESLSIKAS"
/complement (6423. .7718)
/gene="ntrcC"
/note="synonym: Atu3906"
complement (6423. .7718)
/gene="ntrcC"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase
protein [ntrcC]"
/protein_id="AAL44714.1"
/db_xref="GI:17742345"
/translation="MAPEKTAGSKGBATQMDAPALVGSDRKILRAGFIPLVAVS
LIIAAAFGPADREGLILDLVDVSMANVRRLAFQDLAHMALSMPVSMGLGSP
SPITTPSISGCGNAITLSTRLPARKMLGLSETAGALANALKILVIDDMARGE
PMIAGTVPSSHNEPERVYLAAGGIDHDVTLVVPPTSDDLAAAGIDPGVGA
PNIVIAERGVGRIVAKODIMPABEKVIGMPPEWASQGTGSLTLRLDAAASC
DLADNDALSGALADPRYIGAPOSTIRIVLAGFSTDSQGNRVIEKFTFHGDHANY
PQSGSLWIVSQIRWQAELESETGVNAALSAVRPDIYRALDGDGAPGADIRIGQ
DEGDRVDGFVPDPADIVAGVNSFAVTSVPASPSGEA"
/complement (7754. .8341)
/gene="nast"
/note="synonym: Atu3907"
complement (7754. .8341)
/gene="nast"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="two component response regulator"
/protein_id="AAL44715.1"
/db_xref="GI:17742346"
/translation="MTLRDLTILVIDENAIRASIIREGREAGORYTVIHEVNGAR
TTEILOPDIYFDLBNPDMDELFLPOLTNRVGRPIAMVDRDPTASIEAAVYAGISA
YVVDGLKGRVKTILDMASRVAFSRLQRELDASALERRKVIIRAKGITLMKQGL
SESEAFALIRQTMANNEKKISETIAQSVTTAAAGL"
/complement (8443. .9138)
/gene="Acu3908"
complement (8443. .9138)
/gene="Acu3908"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, GntR family"
/protein_id="AAL44716.1"
/db_xref="GI:17742347"
/translation="MLHCSERSAMTETIEGLPLADRLAKGIRTLISGELAQGIS
EAPFSERFDVSRNSLRFAFLTKDGLLRHEANRGVFAVPTMSSVIDIYRVRMVEC

Query Match 9.6%; Score 133; DB 1; Length 10707;
Best Local Similarity 50.2%; Pred. No. 1,9e-15;
Matches 572; Conservative 0; Mismatches 495; Indels 72; Gaps 7;
GALRQMHKHPAIKRAVEDAKIRREADNIGVGSANKFHAAVVDLDSARISDF
YERIAAEHLRCFGLDPEQLHSPVDNMKSLISLMEBGAQAGNEMELYNLSERT
QY 87 CGATGTCGGGGCAGTTGGCAAGCTGGAAGAAATTGAAATTCGCTTCAACGCT 146
DB 5941 CGCTTTCGGCCACCGCAGAACCGGAAGTGAAGGGTGCAAACTCGGTTTCACTCGCGCT 5882
QY 147 CACCGACATGCGCCCGCTGGCGGTGGCCCGGCAAAAAGCTTCTTGAGAGACGAGGCGCT 206
DB 5881 CACCGACGCGCCCGCTCACTTCGCGCGGCAAGGGAATTTCGCCAAGCATGGCAT 5822
QY 207 GTTCGTGCAACTGGAAGCG---CAGGCGCACTGGAAGTGGTGAATGAGTATGGGTCGGA 263
DB 5821 GCTTGAGGTGAGAGTTCTGAAACAGGCTCTTGAGGGGCGACCGCGACATCTCGTGTCT 5762
QY 264 TGGCGAAGTGAAGCGCTCGCAATGCTGGCGCGCGCGCTTGAAGCGCCAGCGTTGGCTT 323
DB 5761 CGCGCG---CGCTCGAAGCGCATGATGGCGGCGCATATCTGACCGCATGCCCTTA 5708
QY 324 CGGCAACGAGCGCGATATGAGGTGCCCTTCAAGATGGGCTTCAACGCGCAACGCGATTAC 383
DB 5707 TCTCATGCAATACCGGCAAGGTGACGCAAGCAACGTCGCGTCCGATGCGATCTCCGCG 5648
QY 384 GGTGCAATGAATCTGGCATGATGAGAACCGGAACATACCGCTGGAAGCGGTAAC 443
DB 5647 AGCGCTTAATCTGACAGCGCAAGCGCAT-----TTCCGTGCGCAAGGAATATGACAGAC 5594
QY 444 GGTGCAATCGCATCAAGGCAAGATTATGTAACCCGCTGTCGAAATAATCAACCGCAAG 503
DB 5593 CGCGCTGCAAGCTGATGATGCTCCAAAGCTGAAGCGGCTTGAAGAAAAGAGCGCG 5534
QY 504 CAAAGCGTTCAATATGAGCGATGACTTCCGCGCGGATCTCAACAATCAACTGCGTTA 563
DB 5533 CAAAGAAATCAAGAGCGCGCATGACTTCCCGCGCGCAAGATCTCGATCCGCTTA 5474
QY 564 CTGGCTGGCGGCGGGGATGCAATCTGGCTCTAATGGCGCGCGCGGCAAGATCTCG 623
DB 5473 TTGGCTGGCTGCTGGGCGCATGCTCGGCAAG----- 5440
QY 624 CCAATCGCGCGCAAGCGCTTGTTCGTGACCCGCGCGCAATATGCTGCAAGCT 683
DB 5439 -----GAGTCTCCACCATGCTGCTTCCGCGCGCAATGCTGCAAGCAT 5393
QY 684 GGAAGCGCGCAATTTTGGTTATTTGCTGCGCGAGCGGTGAACCAAGCGCGTGT 743
DB 5392 GAAAGTCGCGCAATGAGAGTCTTCTGTGTAAGCGAGCGGTGAACCAAGCGCTGTTAA 5333
QY 744 CAAAGCATATGAGGCTCGGTATCAACCATGAAGACTCTGGAAGCAACCGCGA 803
DB 5332 TCAAGGCTATGCTTAACTGCTGTACCAACCGGGAATCTTGAAGAGGCGATCGGAAAA 5273
QY 804 AGTCTTCGCGGTACCAAAATATGAGCGGAGAAATACCCCAACCTATCTGCGGTGAC 863
DB 5272 GGGGCTCGCGATCGCGCGCATGCGGTGAAAAAGAACCCCAAGCGCAAGGCTTTGCT 5213
QY 864 CAAAGGCTGATTCGGGCGCGCATGCTGCTGACCGCGCAATTAACAGAACCGCAAG 923
DB 5212 GATGCGCGTATGAGAGCGCGAATGTCGACGAGATGCGCAACAGAGAGAGATGTC 5153
QY 924 AGCATATGAAATCTGCGCGCAAAACAAATACGTGCTGACGTGAGTGTGCGCGC 983
DB 5152 CACCATC-----CTGCGCAAGCGCAAGGTTCACAGTCCCGCAAGAGATGTTCTCGG 5099
QY 984 TAGCATGAAGCGCACTTTCGATACGAAAGAAAGAGATTAACGCGCGCTACCGGACTT--- 1040
DB 5098 CGGCTTCAAGGCGCAATTAATTAACGCAAGCGCGGCTGTAAGAAATATGCGCTGCA 5039
QY 1041 CAACACCTTCTTTCGCGCAAGCGCGCATATCTGCTTCAAGCATGCGAGTCTGATCT 1100

```
Db      5038 GATGAAGTTCTGGCAGGAGACCAGCTCTCTATCGGTTCCACAGCCATGACAGCTGTTTCAT 4979
Qy      1101 GACCCGACGTGAGGCGCTGGGGGCATGATCAATGATTTCAAACCGGACACTGGTATCTGGA 1160
Db      4978 CACCGGAAAACATCCGCTGGGGCA-----AATTCGCGCCCGGATACGGAAGCTAAAGGC 4928
Qy      1161 TACCGGCAAGAACGTCTACCGCGGACATCTATCTCGCGCGGCGCAAGGAACTGATGCG 1219
Db      4927 GCTGCTTCGAGAGCTGAACCGCAGATATCTGCGCGGACGCGCGCAAGGATCTCGCGC 4869

RESULT 9
LOCUS      Ap003000                      348077 bp    DNA        linear    BCT 15-MAY-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 7/21.
ACCESSION  AP003000 BA000012
VERSION    AP003000.2 GI:14023048
KEYWORDS
SOURCE
ORGANISM   Mesorhizobium loti
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE
AUTHORS    Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Ideasa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
            Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
JOURNAL    DNA Res. 7 (6), 331-338 (2000)
MEDLINE    21082930
PUBMED     11214968
            2 (bases 1 to 348077)
REFERENCE
AUTHORS    Kaneko,T.
TITLE      Direct Submission
            Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel.81-438-52-3935(ex.2338), Fax:81-438-52-3934)
            On May 11, 2001 this sequence version replaced gi:11994975.
COMMENT
FEATURES
            Location/Qualifiers
                1..348077
                /organism="Mesorhizobium loti"
                /mol_type="genomic DNA"
                /strain="MAFF303099"
                /db_xref="taxon:381"
                /complement(169..1065)
                /gene="ml12551"
                /complement(169..1065)
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="BAB49656.1"
                /db_xref="GI:14023049"
                /translation="MALDWDKLVFPAALAEAGSFTFAAETLHLSQSAISROYSALEHD
                VVPLFNRARGVLTEOGEMLFTADVIMLEKLETKSRLETKDRSPGLAVTTVG
                LGAGWLTREVQETLEPEISQLIANEELRLRQADCAIRLQPOOPDLIORLF
                TVHFLYAAPSVYAAHGKPAISVSELRNRIYFGAPVPSHSELNMLETVCPFBGGOR
                VSLQINDLSTIKRAVQGGAGIAMPDVIYNDSGLVQLPTEVPSFTYFAYPDAM
                KNOAKLHVRDFTIAKARSWSF"
                /complement(1156..2130)
                /gene="ml12552"
                /complement(1156..2130)
                /gene="ml12552"
                /codon_start=1
                /transl_table=11
                /product="thioredoxin reductase"
                /protein_id="BAB49657.1"
                /db_xref="GI:14023050"

            1..348077
            /organism="Mesorhizobium loti"
            /mol_type="genomic DNA"
            /strain="MAFF303099"
            /db_xref="taxon:381"
            /complement(169..1065)
            /gene="ml12551"
            /complement(169..1065)
            /codon_start=1
            /transl_table=11
            /product="probable transcriptional regulator"
            /protein_id="BAB49656.1"
            /db_xref="GI:14023049"
            /translation="MALDWDKLVFPAALAEAGSFTFAAETLHLSQSAISROYSALEHD
            VVPLFNRARGVLTEOGEMLFTADVIMLEKLETKSRLETKDRSPGLAVTTVG
            LGAGWLTREVQETLEPEISQLIANEELRLRQADCAIRLQPOOPDLIORLF
            TVHFLYAAPSVYAAHGKPAISVSELRNRIYFGAPVPSHSELNMLETVCPFBGGOR
            VSLQINDLSTIKRAVQGGAGIAMPDVIYNDSGLVQLPTEVPSFTYFAYPDAM
            KNOAKLHVRDFTIAKARSWSF"
            /complement(1156..2130)
            /gene="ml12552"
            /complement(1156..2130)
            /gene="ml12552"
            /codon_start=1
            /transl_table=11
            /product="thioredoxin reductase"
            /protein_id="BAB49657.1"
            /db_xref="GI:14023050"

            1..348077
            /organism="Mesorhizobium loti"
            /mol_type="genomic DNA"
            /strain="MAFF303099"
            /db_xref="taxon:381"
            /complement(169..1065)
            /gene="ml12551"
            /complement(169..1065)
            /codon_start=1
            /transl_table=11
            /product="probable phosphohexose isomerase"
            /protein_id="BAB49661.1"
            /db_xref="GI:14023054"
            /translation="MSLNDYLVRSAAITMVERDLTGEMERASAVVTLISGKAL
            LIANGGSASDAIHITAGELVGRFLKERAIVIALPANAATITANGNDYGPPTVSRO
            VEAHGTAGCVLAISTGNSPSITAAQAQRMDMTVIGLTGDTGKKPLCDITLLNV
            PSTSTPIIOOGLCLYHHLCEVEARLSNG"
            /complement(5129..6619)
            /gene="ml12562"
            /complement(5129..6619)
            /gene="ml12562"
            /codon_start=1
            /transl_table=11
            /product="ADP-heptose synthase"
            /protein_id="BAB49662.1"
            /db_xref="GI:14023055"
            /translation="MTKNPPSPPEPLHRAIRFGGVTVLVGDLILDRFVNGVIRIS
            PEAPLVLAGKETSAMGAGNVANIVISGARALPVSIVGDTGADSLVRLAALGA
            ETAGLSQGRMTSSKSRFSALNOQVLFDEEIVLDETETAGLIRHFRALAGAEI
            VILSYGKGLLDGVAELIATCREAGRPVAVPGRIYARAGATATLPKKEGDEA
            VGHAFADDEIVAAARELISAHGPFVAVATSEKMSVVGDEAHHTINTQAREVDVS
            GAGTIVATFPALALASGADPVAAASINAAAGVYVKGRTARLTTELTGALFRSHGP
```


PUBMED 12597275
REFERENCE 2
AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,
Sasamoto, S., Matsumoto, A., Iwasawa, K., Iriyuchi, M., Kawashima, K.,
Kohara, M., Matsumoto, M., Shimizu, S., Tsurumoto, H., Wada, T., Yamada, M.
and Tabata, S.
TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL DNA Res 9 (6), 225-256 (2002)
MEDLINE 22485002
PUBMED 12597279
REFERENCE 3 (bases 1 to 300150)
AUTHORS Kaneko, T.
TITLE Direct Subinjection
JOURNAL Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; 2-6-7
Kazusa-Kametari, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/rhizobase/,
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
FEATURES
source Location/Qualifiers
1. 300150
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
/complement(14..862)
/gene="b112729"
/complement(14..862)
/note="ORF ID: b112729"
/protein_id="b112729"
/codon_start=1
/transl_table=11
/protein_id="BAC47994.1"
/db_xref="GI:27350986"
/translation="MDNRERRMTPRQIIGKIGLMISFVIVSPATLPIAMLSLTK
FEIDNAYPPVPEPHIAKMYADVANSRPLTYFINSLVGSGATATALLIVGVPAGY
GIAMAAHKSATVILIAKRTTGGSTYIPIFLPQMGILGTVIPQIIHLVTVPIVI
WMIGYFETTPLEBAPALIDGKTRMQVFRHVALPIAKPGIAPAFILAVIFSMNPF
GIVLAGRETRTLPVAVYNNISFDQLSMGLAALIVTLPIVLLTFAPQRIVAGLTA
GAVKGG"
/complement(835..1773)
/gene="b112730"
/complement(835..1773)
/gene="b112730"
/note="ORF ID: b112730"
/protein_id="b112730"
/codon_start=1
/transl_table=11
/protein_id="BAC47995.1"
/db_xref="GI:27350987"
/translation="MSALTQADPAAARSDAPEKELRPSYPFVVPALVVIATIIIF
PWFETIMSLNEMKSPPTFVGLSNYLRLTSDPFIKAVGHTMYTALSVLLPVLG
TFAAVFVHOKFAGRGFLRGVFLPMMAFVALVATWMEFHOLGIVNTIISLVGPA
QIMVHPATVPSIVLVETWMTPLVMLIVGLAIPREPSAIOIGANIMQMRP
ITLPIPMFLPTAGMIRNIDAVKSPDITFATIGGPGSASETINIYLSVARTTYDLG
YSSAIVAVFELIIVLAAVMLYAKQMWTEIASDA"
/complement(1770..3152)
/gene="b112731"
/complement(1770..3152)
/note="ORF ID: b112731"
/protein_id="b112731"
/codon_start=1
/transl_table=11
/protein_id="BAC47996.1"
/db_xref="GI:27350988"
/translation="WAKSLVSLPKTMSHREDPEVVDYRISRTLLTGTAAGALSITG
LFAABVMWKYAGTKLEVTILAKPGNDLQNIKEFELTGIQVEABQIPQOQROK
VIVFTSGRPSDVHLSTYVQKQFQEKAGMLADMTPEKQPSLTAIPDLVBSDDPQOK
LOYAKNDKGMSTLPSMSVDYFLTYNKELFQKGVAVPFTLDEMVAALAKITDPREGT
YGFVGRGLNANMALMTNPFVNYGSEPLDDKKNILLDGEALIAATKIVQTLTKKAP

gene
CDS
/complement(3179..4012)
/gene="b112732"
/complement(3179..4012)
/gene="b112732"
/note="ORF ID: b112732"
/protein_id="b112732"
/codon_start=1
/transl_table=11
/protein_id="BAC47997.1"
/db_xref="GI:27350989"
/translation="MDRORLTVHRLGIDHGMKTRRAIVATTLAAAPALLAGEIAP
EARRSGYAPWGPDTTAMODDTANPQMFLVLDGELMTKRTSAAEAKADCHDADGS
MGVARYAPAEKAGRPITLQRIINLCANQOARPLPYEGRDLTASAFPAHQSG
VAITAGDDPOLKQFVQGRDLFMOREGOLNACTNCHDNFRLHAGAITAOQPFY
PLYRLQWGLSLERLRSCMGVAAQANDYGAPELVALLVLMGRARGLPHETPAVR
P"
/complement(3955..4272)
/gene="b112733"
/complement(3955..4272)
/gene="b112733"
/note="ORF ID: b112733"
/protein_id="b112733"
/codon_start=1
/transl_table=11
/protein_id="BAC47998.1"
/db_xref="GI:27350990"
/translation="MAALINVPKARKRGDVIERITLTSHIMETGRHRTADGLVPRDI
ITSFICRNGTGVPRADLFPALANPYLSFVVAESGCFEFTWGDNGYSTASASI
TVE"
/complement(4272..4739)
/gene="b112734"
/complement(4272..4739)
/gene="b112734"
/note="ORF ID: b112734"
/protein_id="b112734"
/codon_start=1
/transl_table=11
/protein_id="BAC47999.1"
/db_xref="GI:27350991"
/translation="MPTTRROFLSAGGITAAGTPIYTLRPLENTPAMLTATIRNV
GSAPLRIGKVDKIDPPLVENGTVPWITSVSPMTDVIYSIHVFNENKPNQNGNF
HLGPVAGRAQVSTRIRLADTKQVAILRSDSFQVAVAVVTLLAACTEKN"
/complement(5385..6665)
/gene="b112735"
/complement(5385..6665)
/gene="b112735"
/note="ORF ID: b112735"
/protein_id="b112735"
/codon_start=1
/transl_table=11
/product="flavocytocrome c flavoprotein subunit"
/protein_id="BAC48000.1"
/db_xref="GI:27350992"
/translation="WRPNAPVTRRNVLGITAAATSVAVSILRAQSGARVVVGGG
FEGAACAPALKRANKLOVTLIEPKVPISEFSPNSVAGREIEAQOFGYDKVAEG
ITVVAQATTVDPERRSVTTDGAITLVDRLVLSGIDFHALGDEASSEKMPHA
WKAGAQTLLRLROLEMDDGGVATAIPANBSRCPAPYERASLILAHYLTKEKPRAKY
LILDAKDSFSOORLEPERAKELXGMIRERIGLSOGGRVTSVDASTKTIYTERGNTPD
VANVTIPQORAGHLAIDAGADATGWCPIIDPTFESKLVNPHVITGDACIGGQIPAS
AASAGKCAACAAIVNLLAGRAPETRLGVCTTVAPQVGSAGNYOPRGDIYAE
GGATSPVAPRELRAREADEARMEQITTAADTFG"
/complement(6782..8968)
/gene="b112736"
/complement(6782..8968)
/gene="b112736"
/note="ORF ID: b112736"
/protein_id="BAC48001.1"
/translation="putative aldehyde dehydrogenase protein"
/protein_id="BAC48001.1"

RESULT 13
LOCUS AP005951/c 298700 bp DNA linear BCT 12-SEP-2003
DEFINITION Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 17/31.
ACCESSION AP005951 BA000040
VERSION AP005951.1 GI:27352606
KEYWORDS Bradyrhizobium japonicum USDA 110
SOURCE Bradyrhizobium japonicum USDA 110
ORGANISM Bradyrhizobium japonicum USDA 110
REFERENCE 1
AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idegawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsunoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110
JOURNAL DNA Res. 9 (6), 189-197 (2002)
MEDLINE 22484998
PubMed 12597275
REFERENCE 2
AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idegawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsunoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL DNA Res. 9 (6), 225-256 (2002)
MEDLINE 22485002
PubMed 12597279
REFERENCE 3 (bases 1 to 298700)
AUTHORS Kaneko, T.
TITLE Direct Submission
SUBMITTED (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7 Kazusa-kametari, Kisarazu, Chiba 292-0812, Japan (E-mail:kaneko@kazusa.or.jp, URL:htcjp://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935 (ex.2338), Fax:81-438-52-3934) Location/Qualifiers
FEATURES
source 1. 298700
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
48. 512
/gene="b1r4344"
48. 512
/gene="b1r4344"
/note="ORF ID:b1r4344
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49609.1"
/db_xref="GI:27352607"
/translation="MTQATPASTTPKRRPRPQVNLTDAAQRISLTLQADSEIVG
LRGVNKGCGACGQSYVEYAHDPVPTDEVEDKGVILVDPKAVFLILGHRDGLGRO
DAGPVRLQOPGGLRLRLRRIGRAAAGEBDRVGPALPFRSAVVLAKARTHTA"
584. 1069
/gene="b1r4345"
584. 1069
/gene="b1r4345"
/note="ORF ID:b1r4345
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49610.1"
/db_xref="GI:27352608"
/translation="MGQFRODDDEBRCAMREPLILDFADGPTVYIRKMFSGYISA
DGTNPLALRLAGLFLRADELTLDPETESGKFOYSTRAKTYTVNSVLPALPDDDS
QELAQMARALAAAGRAKYKQRPKGRKASKTYAPKPKPARKPAKAVRVRKAGAH

gene
CDS
R
complement(1134..1664)
/gene="fabA"
complement(1134..1664)
/gene="fabA"
/EC_number="4.2.1.60"
/note="ORF ID:b114346"
/codon_start=1
/transl_table=11
/product="3-hydroxydecanoyl-[acyl-carrier-protein]
dehydratase"
/protein_id="BAC49611.1"
/db_xref="GI:27352609"
/translation="MPNPHDFPTPPSYRRELLASSSEGVFGPQNAOLPAPNLMMD
RISBISLDGGERGKHIVGELDIVGHHFFDCHFGDPLMPSSLGLDAWMQITGLG
WSGSPKGRALIVGEVCELTPTVORVRYSVAMRWVRGLVIGIADGRVADGAC
VFTAKDMRGVLTAKAVE"
complement(1763..2830)
/gene="b114347"
complement(1763..2830)
/gene="b114347"
/note="ORF ID:b114347
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49612.1"
/db_xref="GI:27352610"
/translation="WVKVLDHERHTMAFAEVALGOIRSLQTAIPRNYEIWYVATGY
NAPLNKININETLARKKLTLEADLEQIYEYVLSHITTDRIIDKARVIGEIDVWKL
GPAIWMGYSYDASLSGATERLSAKSREQIKAVETLVSTSEMEETNALDRILTS
RNEISNLOOSLEAIRASLTDPILGINKRMDMIGMAVOSLAGSEPLILPID
HFKSPNDSYGHITGQVRLVYLSMKQRTIKODITARVGESEFVILLPTALRQALT
ADHRRVAKELKKKSTGELIGRVTVISVSMKRPDDTDLITERADCLYAAKNG
RNRVICADPEFAVETHSRVA"
complement(2994..4496)
/gene="rhlE"
complement(2994..4496)
/gene="rhlE"
/note="ORF ID:b114348"
/codon_start=1
/transl_table=11
/product="ATP-dependent RNA helicase"
/protein_id="BAC49613.1"
/db_xref="GI:27352611"
/translation="MSFSNLGSEKVLAAVAATGTTPTPIEOAIPIYLAKQVIGI
AQTGKXTAAFLVLPMLTILEKGRARARPRLLLEPTRELAQVKEENFVRGAGQGLN
VALLIGVSPGQDALKTRGVDLIATPGLDHTERGLLITVELVITDADMLD
MGFIPDIERICKLVEPTROTLEFATMPPEIRITETELHNPQKVEVSKPATVAVT
OSQVPAKKAHKEKRLRLRLBEAKDKNAIIFCNRKRVATVHKSLOKSPGVALH
GMDQPARMALEQFRKSELPLVASDVAAKGLDIPYSHVYNPVPHHPDDYVRVG
RTGRAGSGTALSTVPLDQKSWAIEKLIQGITRABGDYVHAHAGADATDPREOR
GNERSRGGRKQGRDRDRSHRPEARPSAARPAPEARPREPRPPREARSPPEPR
HGSRRQANASHPVSGRPEPRRADTEPCDHSILPAFLRPVSPAGA"
complement(4787..5782)
/gene="b114349"
complement(4787..5782)
/gene="b114349"
/note="B114349 family
ORF ID:b114349"
/codon_start=1
/transl_table=11
/product="transcriptional regulatory protein"
/protein_id="BAC49614.1"
/db_xref="GI:27352612"
/translation="MYRWCTDVEVEQERFDYREYRAGLFGVTABLERRRADPFGE
PSLRQAGGLVELKASPYTVRSISDIATPGDSICVYQOISGSGWGMGMSGPAI
ANGSRATSHTDLPYRTAPLDAGGPHRLIKIPVSIIPQDRIIRLITKTPDHALAP
LIAACFADLGEVAAAGDGDGADATSLVQTSIVQALAHALITERGIVRGSRLQ
AALRTARLSQARLILARQLONPNDLPAWADLIGSVRLHMLFPAEKSFSQYVDE
RLKRSRLRLREPERLIDIAVTSCEFSLATYTVRVNAAYGMAPDFFARVDEGH"
5935. 7503
/gene="b1r4350"
5935. 7503


```

/gene="b1r4350"
/nc="ORF ID:b1r4350
/hypothetical protein"
/codon_start=1
/translate=11
/protein_id="BAC49618.1"
/db_xref="GI:27352613"
/translation="MDCTRGAGGKPSGNGNFSMTBRLFRILVTGLAASLGAFALPAC
AERKVALVGNNDYRVPKLLKAVNDARTMGDTLKQGFVWVAENQSRQFSEETLLA
PRAIEPBDTAPFPVAGHGEIAGQVLLPTDPAATBQEEELVDRASILADRIVERL
ONKARTSILIFPACRNPENPERKGTAVAGGGLAAMMOLBEVGFVSFASPRQOTLD
RISNDADPNVSVPRTFPAKELLQGENIUNOVAORTRIYSEMAUTYKHQVPTPYPM
VNDVPLSGAKADAAAPADPPPOKLAALPPVSPRPVKESTTAAPIASFSEHNGMSV
VSEFADPTGISHRMAGNDQFRETGFITLIDPRTKRPMPSTLEPPDAQGITIEVRY
VQSGMDGPFPIRFPDEALIRDKRIIDMTATMSLSPFEFNGLLVYTHLVSYSKA
IEEVRIGITAVPNQVLRKMGCDMDPSAITAGMPLMKLAIPATQFVSVELTYRDSV
SEKSFRAARSN"
7987.9162
/gene="b1r4351"
7987.9162
/nc="ORF ID:b1r4351"
/translate=11
/codon_start=1
/product="FAD-dependent monooxygenase"
/protein_id="BAC49616.1"
/db_xref="GI:27352614"
/translation="MMLGYLGRAGIDVVLEKHADFPDRDGTVPHTLQVMEIG
LIDPLKPHORLOKMDLFGTTPVRIADLSRLNTKYPPIAFMPQMDPLNPLFEAKGR
PASIRYMNTEAVDLIRGRTPAQVAKTPEGIYDIEADLTICDRHSVRRACILE
VBEIGAPMDVLPFRAGKADETSVPARVEPQMTTFDGDYMQCAVYAKQHDVA
KARGQALLDDVVRKAPILRSGLADVSFDIILTLVAINRLARWTRPGLLIGDAH
AMS PVGSGVNLAVQDAVATNLADKLQHCPSDELDVARRRREFPVMTQAMQVI
VNNTISGALQSGDRPLKVLVRLITLALPMLQGI PARLIALGVREHVSKAPAS"
9644.10108
/gene="b1r4352"
9644.10108
/nc="ORF ID:b1r4352"
/hypothetical protein"
/codon_start=1
/translate=11
/protein_id="BAC49617.1"
/db_xref="GI:27352615"
/translation="MFGMEGTISGLDNKGSPLNTAFAGADVFSRANVLAIVAGAG
PBYVNNLPIYKGGVAGVNNRIASVTSVPATSGGOTTHANNGTVAAGHEVGTNRMI
VLEIYVYAFGSGTYQLGTTTNYTFDAPRDLQMAVRASTYFADAPVARY"
complement(10205.10975)
/gene="kxud"
complement(10205.10975)
/nc="ORF ID:b1r4353"
/codon_start=1
/translate=11
/product="2-deoxy-D-glucuronate 3-dehydrogenase"
/protein_id="BAC49618.1"

```

Query Match 8.1%; Score 113.2; DB 1; Length 298700;
 Best Local Similarity 54.9%; Pred. No. 1.2e-11;
 Matches 267; Conservative 0; Mismatches 213; Indels 6; Gaps 2;

```

QY 131 TCGGCTTCATCAAGCTCACGACATGCGCGCGTGGCGCGCGCGGAAAAAGGCTTCT 190
    |||||
DB 263952 TCGGCTTCATTCGCGTTCGATGCGCGCGCGCTGATCGTCCGCTGCAAGAGGCTTTA 263893
    |||||
QY 191 TCGAGAGCAAGGCGCTGTTCTGTCACACTGGAAGCGAGCGCAACTGGAGAGTGTATGG 250
    |||||
DB 263892 CCGCCGCCCAAGGCTTCGAGCTGGAAGCTGTGCGCAGAGGCTCTCTCCCAAGCTCCGCG 263833
    |||||
QY 251 ATAGGCTCGTAATGGAGCACTGAGCGGCTCGCAATGCTGGCGCGCGCGCTTAGCGCG 310
    |||||
DB 263832 ACAAGCTCAATATCGGCTGTTGACGCGCGCCCACTGCTGCGCGCGCGCGCATGCGGT 263773
    |||||

```

```

QY 311 CCAAGCTGGCTTCG---CACCAAGCCGATATTCAGAGTCCCGTTACAGATGGGCTTCA 367
    |||||
DB 263772 CTTCACTCGGCTCGGCGCAAGTCAGATGCGCGCGCGCTTCAATCTCGGCAATCA 263713
    |||||
QY 368 ACGGCAACCGGATTAAGGTGTCCAAATGAATCTGGCATGAGTGAAGCCGAACATACCGC 427
    |||||
DB 263712 ACGGCAACCGGATTAAGGTGTCCAAATGAATCTGGCATGAGTGAAGCCGAACATACCGC 263653
    |||||
QY 428 TGGAGCGGTAAACCGGTGCATCCGATCAAGGACAGATTATCTMAAACCGGTGTCGAAA 487
    |||||
DB 263652 GGCACCGCTTCGATCCGCTGTTACCGGCGAAGCGGTGCTC---GGTGTGCGCGAGGC 263596
    |||||
QY 488 AATCAAGCCGCAAGGACCGGTTCAATATGCGCATGACCTTCCGCGCGGATCTGACA 547
    |||||
DB 263595 GCGCGAAGCTAGGTGCGCATCCCGCTACCTTCGCGCATGACCTTCCGTTCTCGACCCACA 263536
    |||||
QY 548 ACATCAACTGCTTACTGCGCGCGCGGTATGATCTCGGCTACTATTCGCGCG 607
    |||||
DB 263535 ATTAACCAATTCGCGTCTGATGCGCGCGCGCGCGCGCTGATCCCGACGAGATGTGCGGC 263476
    |||||
QY 608 CGCAGG 613
    |||||
DB 263475 TGGTGG 263470
    |||||

```

RESULT 14
 AE007223/C
 LOCUS
 DEFINITION
 complete plasmid sequence.
 AE007223 AE006469
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Sinorhizobium meliloti 1021
 Sinorhizobium meliloti 1021
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

REFERENCE
 AUTHORS
 Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Adola,A.P.,
 1 (bases 1 to 12129)
 Gurjel,M., Hong,A., Huizer,L., Capela,D., Galibert,F., Gouzy,J.,
 Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
 Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
 Nucleotide sequence and predicted functions of the entire
 Sinorhizobium meliloti pSyma megaplasmid
 Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)

JOURNAL
 MEDLINE
 PUBMED
 21396509
 11481432
 2 (bases 1 to 12129)
 REFERENCE
 AUTHORS
 Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Adola,A.P.,
 Bartley-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
 Gurjel,M., Hong,A., Huizer,L., Hymen,R.W., Kahn,D., Kahn,M.L.,
 Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
 Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
 Direct Submision
 Submitted (29-MAR-2001) Biological Sciences, Stanford University,
 371 Serra Mall, Stanford, CA 94305, USA

FEATURES
 source
 1..12129
 /organism="Sinorhizobium meliloti 1021"
 /mol_type="genomic DNA"
 /strain="1021"
 /db_xref="taxon:266834"
 /plasmid="pSyma"
 complement(103..318)
 /locus_tag="Sma0580"
 complement(103..318)
 /locus_tag="Sma0580"
 /function="Miscellaneous; Unknown"
 /note="glimer prediction"
 /codon_start=1
 /translate=11
 /product="hypothetical protein"

```
gene
CDS
    /protein_id="AAK64962.1"
    /db_xref="GI:14523387"
    /translation="MSSRALQRSYTRGGLGQFWNAKASFRANSDCSSRTACFDGAIK
    ALVGATKPPRSRQTSALQIADGFPVNG"
    complement(394..1194)
    /locus_tag="SMA0581"
    complement(394..1194)
    /locus_tag="SMA0581"
    /function="Cell processes; Transport of small molecules"
    /note="glimmer prediction Excellent match (e-87) to
    nitrate transport protein ATP binding component nasd from
    Klebsiella oxytoca starting at amino acid 35 of glimmer
    prediction. Met33 of glimmer prediction has good SD
    sequence and was called SMA0581 start. Chrd, Nasd and Nrtc
    are synonyms. Possible operon with SMA0581, SMA0583 and
    SMA0585"
    /codon_start=1
    /transl_table=11
    /product="Nitrate transport ATP binding protein, probable"
    /protein_id="AAK64963.1"
    /db_xref="GI:14523388"
    /translation="MTKSYLSLELDKSFERGQRTVLAQVSLTVDKGEFISIGHS
    GGGKSTLLNIVGGLTQATTVGVLLDGKVVDEGPDPAVFNQNSHLPMITVENVLA
    VKVFSRTNKKQERHEWTRNLELVQMAAAKHPSEVSGMKQGVGIALAMPEKV
    ILDEPFGALDLTRAHLQDOVMOIHATGNTVMTHTHDVDAVALLSDRIVMWMTGPS
    AVGRIIDVPLARPRRIELASDRITVLTRESVLAKEVLRHAFVNAE"
    complement(1208..2101)
    /locus_tag="SMA0583"
    complement(1208..2101)
    /locus_tag="SMA0583"
    /function="Cell processes; Transport of small molecules"
    /note="glimmer prediction Possible operon with SMA0581,
    SMA0583 and SMA0585. Excellent match to nitrate transport
    proteins in Klebsiella (NasB) and cyanobacteria (NrtB)"
    /codon_start=1
    /transl_table=11
    /product="NrtB, Nitrate transport permease protein"
    /protein_id="AAK64964.1"
    /db_xref="GI:14523389"
    /translation="MSVTNMLKPKQAPOTPAQVIALGOTASRGIDGRLSRFTQTVT
    NLLPLVLTFTFLAMQLICSSPESLSPPSRVLESWELIAPFIQGVQVQGFWMH
    VASLQRVVALGYAMAAGVVALGTLVGSALMRGADPFOVLRVPLAMVLSIAA
    PDGGPSAIFVFIETAIWPIIINTAVGFINIPODYONVAKVRLNSFEFRGIMLPA
    APYFTGRIIGIGLSMLATVAABMLGIVGICGFITDANWSSLISDIYALTYGVIG
    FLDDLIALVGAATVGTANA"
    complement(2164..3456)
    /locus_tag="SMA0585"
    complement(2164..3456)
    /locus_tag="SMA0585"
    /function="Cell processes; Transport of small molecules"
    /note="glimmer prediction Excellent match (e-143) with
    nitrate binding protein precursor proteins from several
    cyanobacteria. Possible operon with SMA0581, SMA0583 and
    SMA0585"
    /codon_start=1
    /transl_table=11
    /product="NrtA-type periplasmic nitrate transport binding
    protein, probable"
    /protein_id="AAK64965.1"
    /db_xref="GI:14523390"
    /translation="MTKTLFGGLTRRSVLTQTATAWVGAARTLLPSGAFAGAGPRT
    AKATIGFALTDSAPLIITAKEKGLPDKRGMTVEVYVKAQSMGTTTDDNLVLSAGACID
    GAHIIITPMPYLLISTGKVTONQPLPALLIARLNDQALISGAAYADLVKSIDASVLK
    DAFARKKGGEAAKAVMPFGGTHDLIRWMAAGIDPDKQVETIIVPPPMVAMNK
    VGTMDCFVGEPMNQIVQKIGYAVNTAAIIMASHPEKSPFARMADWAKENRAVAL
    VWAIIEAOMCDMANKDEIAKIVGRKSWFNVPDIODRLKGEYDYGKIVENSPH
    FMKPRERIASVPSQSHDAWELTENIRMGSLASDPIKGLIAVYNNEDIWREAKDLGI
    SDIPASTSRGPEPTFDGKVFDPANPEYTLKSLAISRIA"
    4274..4732
    /locus_tag="SMA0590"
    4274..4732
    /locus_tag="SMA0590"
    /function="Miscellaneous; Unknown"
```

```
repeat_region
gene
CDS
    /note="glimmer prediction"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK64966.1"
    /db_xref="GI:14523391"
    /translation="MTCGPYHDSVRFANACACISLSDTGRFGRLPGEAGQVPA
    RHVSPARFGCTDDGSGSGTHAFTPNPRILASQTOADQDSHRQERTDLSRHSRQ
    SGRQRAAFALITAPHSMAARRGRAPVAMAVATVYPRPGVCAIWS"
    4518..4957
    /standard_name="ISRM17"
    /note="SMA002; predicted by homology, possible ISRM17
    fragment, 92% identical over length to ISRM17"
    /evidence="not experimental"
    /tpc_family="ISRM17"
    4761..4955
    /locus_tag="SMA0591"
    4761..4955
    /locus_tag="SMA0591"
    /function="Miscellaneous; Hypothetical/Partial homology"
    /note="glimmer prediction Excellent match to 64 amino
    acids at carboxyl terminus of much larger (44 kDa) protein
    from Sinorhizobium NGR234"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein, fragment"
    /protein_id="AAK64967.1"
    /db_xref="GI:14523392"
    /translation="WVRSRWYRKLRNFRAGIEAGISGLTRTYGLAHTWRGLHHFET
    YVSSVVAVNLALFARLRPT"
    5249..6421
    /locus_tag="SMA0592"
    5249..6421
    /locus_tag="SMA0592"
    /function="Miscellaneous; Hypothetical/Global homology"
    /note="glimmer prediction match (e-19) to two similar size
    (45 kDa) proteins in Sinorhizobium sp. NGR234 over entire
    length"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAK64968.1"
    /db_xref="GI:14523393"
    /translation="WALRSANLTFDETAGTIVETANGSTRPAHSDMNEGIACCF
    STQREHEKVGGLHPEFQHLGPEGMALREOQARSATVEEDIGLLRLYGADCIAGVSR
    PPDDAQLPEITEATVSPEKTVSGVQKLLVTKDDENRFVPSAGSALYIAKFNSSDR
    IDNLVRNELLSLRTAAVIGEREVTFASLTAVIDETALIVTRDRPNGEKRLSD
    CAQILSKRGDYGAGYDAAYEDIAIIRHSSRAPIDLRFENRLIYFTLIGNDAD
    LKNSFLRPTGRLTRSPAYDVNTATPYGFOOTLALSTIGEKIHEAANOALFPAFK
    EIGLPRADIDQTFKQKQVEKASITIRPPDAEPADGFVHRKEIVDNSCLRILET"
    6397..7428
    /locus_tag="SMA0594"
    6397..7428
    /locus_tag="SMA0594"
    /function="Miscellaneous; Unknown"
    /note="glimmer prediction"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK64969.1"
    /db_xref="GI:14523394"
    /translation="WPNNTGNLAPLAKQRYVEBALRRRKASGLTQKHSALAGVSHPT
    MAAFERGETTLTALAALDILRVGLVDPRTGDDTARVRAAFERWRNLVVALPDSP
    ARFPMGWRFYWTMGEDLAKSELTFEEILKAVVRKMGWPFMLPTREALIOPRVDG
    LIECLIAQGEVEVERGFNDPAPCDWRAPASGRMLIIGYODGAGETEPAGIILDTTL
    PLMRNGEVLIAEKIASILRDADQATVAVHPRAMPTGIRGVLRSEWNPISGLIYVEGH
    GARSDEAVLEAKFSANDIESRLAECMLPLTISLYERFVAGLSLNRVASEVORLNSP
    ISKERPRR"
    7698..7919
    /locus_tag="SMA0596"
    7698..7919
    /locus_tag="SMA0596"
    /function="Miscellaneous; Unknown"
```



```

/note="glimmer prediction"
/codon_start=1
/translate_table=11
/product="hypothetical protein"
/protein_id="AAK64970.1"
/db_xref="GI:14523395"
/translation="MCTTKITKRLFLAMWQFASVTNNMLSLSRLSCRIQLIFQINDF
DIEFATPDEGADIALAVNPFKLFKREG"

Query Match      7.9%; Score 109.6; DB 1; Length 12129;
Best Local Similarity 52.3%; Pred. No. 6,1e-11;
Matches 402; Conservative 0; Mismatches 309; Indels 57; Gaps 5;

QY 454 ATCAAGCGAGATTATCTAAACCGGTCGTGAAATAATCAAAAGCCGAAGCCGCGTTTC 513
DB 2997 ATGAGCGCCCTCGGTGCTGAAGAAGATGCTTCGCCAAGAAGAGCCGGTGGCGAGCGCC 2938
QY 514 AATATGGCGATGACCTTCGCGCGCGGATCTCAACAACATCAACCTGCTTACTGGCTGGCG 573
DB 2937 AAGGTGCGCATGACCTTCGCGCGCGGATCAACGACCTCTGGAATCCGCTATTGGCTGGCC 2878
QY 574 GCTGCGCGATTCATCTGCTGCTACTATTCCGCGCGCGGACCAATTCGCGCAATCGGC 633
DB 2877 GCGCGCGCGATCAACCGCGCAAGATGTCGAACCATCTGCT----- 2834
QY 634 GCAGACGCCCTTGTGTGTGTGACCCCGCCCGCGAAATGCCGTCACGCTGGAAGCCGCG 693
DB 2833 -----CCGCGCGCGGAGATGTGTGCGCAACATGAAGTGGCG 2797
QY 694 ACCATTTTCGATTATTCGCTGCGCGAGCCGTGGAACGACGAGCCGCTTTCAGAGGCATA 753
DB 2796 ACCATGACCTGTTCTGCGCTGTGCGGACCACTCTGGAACGACGACCTCTCAACGAGATC 2737
QY 754 GCGGTGCGCGATTCACCGATGAAGAACTGTGAAGGACACGCGGAAAAAGCTTTCGCG 813
DB 2736 GGTTCACACGCTGTCAACCGCGCGAGATGTGGCCCGACGATCCGGAAGTCTTTCGCC 2677
QY 814 GTGACCAAAACAATGCGCGGAGAAATACCCCAACAACCTATCTGCGGTGACCAAGCGCTG 873
DB 2676 ATGCGCGCGGACCTGGGTGCGAAGAACCGCGCGCGCTCAAGCGCTGTCAATGGCCATC 2617
QY 874 ATTGCGCGCGGATCTGCTGCGACCGCGCAATPAACAAGACCGCAAGAACCTATCGAA 933
DB 2616 GAGGAAGCCGCGAGGTGTGCGACG-----ATATGGCGCAACAAGACGAATCTGCCAAG 2563
QY 934 ATGCTGCGCGAATAACAATAGTCGTGCTGACGTGGAATGCTGGCGGCTACATGAAC 993
DB 2562 ATGCTGCGCGAAGCGACGCTGCTTCAACGTGCCCAAGAACATCTGATCGGCTGAAA 2503
QY 994 GGCACCTTTCGAATACGAAAAAGACGATPAACGCGCGCT--ACCGGACCTCAACACCTTC 1050
DB 2502 GGTGAATACGATCAACGCGCAACGCGAAGATGTGGAAGAACGCCCACTTCATGAATTC 2443
QY 1051 TTTCGCCACGCGCGCCAGCTATCCGTCTTACAGACAGTGCAGTGTGATTCGACCCAGCTG 1110
DB 2442 TGGCGCGAACAACCGCTCTTACCGCTTCAAGAGCCATGACCGCTGTCTTCCACGAGAT 2383
QY 1111 AGCGCGTGGGCGCATGATGATGAATCAACCGGACCAACTGTATCTGATACCGCAAG 1170
DB 2382 ATCCGCTGGGCGCAAGCT--TGATCGGAATACGACATCAAGGCGCTGATCGCC----- 2332
QY 1171 AACGTACCGCGCGGACATCTATCTGCGCGCGCAAGGACCTGCTC 1218
DB 2331 AAGGTCAACCGTGAAGACATCTGCGCGCAAGCGCAAGAACCTCGCG 2284

```

```

RESULT 15
LOCUS      AP006575      299750 bp      DNA      linear      BCT 18-OCT-2003
DEFINITION Gloeobacter violaceus PCC 7421 DNA, complete genome, section 8/16.
ACCESSION  AP006575 BA000045
VERSION     AP006575.1 GI:35212526
KEYWORDS

```

```

SOURCE      Gloeobacter violaceus PCC 7421
ORGANISM    Gloeobacter violaceus PCC 7421
REFERENCE   1
AUTHORS     Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H.,
            Tsuchiya, T., Sasamoto, S., Watanabe, A., Kawashima, K., Kishida, Y.,
            Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N.,
            Shimo, S., Takeuchi, C., Yamada, M. and Tabata, S.
            Complete genome structure of Gloeobacter violaceus PCC 7421, a
            cyanobacterium that lacks thylakoids
            DNA Res. 10, 137-145 (2003)
REFERENCE   2
AUTHORS     Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H.,
            Tsuchiya, T., Sasamoto, S., Watanabe, A., Kawashima, K., Kishida, Y.,
            Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N.,
            Shimo, S., Takeuchi, C., Yamada, M. and Tabata, S.
            Complete genome structure of Gloeobacter violaceus PCC 7421, a
            cyanobacterium that lacks thylakoids (supplement)
            DNA Res. 10, 181-201 (2003)
REFERENCE   3
AUTHORS     Kaneko, T.
            Direct Submission
            Submitted (15-AUG-2003) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research, 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 252-0818, Japan
            (E-mail: kaneko@kazusa.or.jp, URL: http://www.kazusa.or.jp/cyano/,
            Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
            Location/Qualifiers
                1. 299750
                    /organism="Gloeobacter violaceus PCC 7421"
                    /mol_type="genomic DNA"
                    /strain="PCC 7421"
                    /db_xref="taxon:251221"
                    /complement(199..606)
                    /gene="g111960"
                    /complement(199..606)
                    /note="ORF ID:g111960
                    hypothetical protein"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="BAC89901.1"
                    /db_xref="GI:35212527"
                    /translation="MAQKSRBALRVEIYICPTGVALRMPPGKHFFVSTTVOAREL
                    LGLTKKTATLHAATTTIVDRTRIEPFPDNGAAMIACHRKSGQVSQVAATGEQW
                    QQRTGVDRRRRPSVGEFTYRYSRIRLRTHG"
                    /complement(797..1249)
                    /gene="g111961"
                    /complement(797..1249)
                    /note="ORF ID:g111961
                    hypothetical protein"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="BAC89902.1"
                    /db_xref="GI:35212528"
                    /translation="MKTVTLLIPVYTHIDFVGHVSNIYIOMEIIGRMKMEIGMP
                    VAKTSQGIAPULIETRIYKISIKGDCAHNEIWDDELKSVSNVAAPKRCNGEGLV
                    ATDHQGVFVDVQKMPRLPAEMERFEPYLOEAPSETRVARSRK"
                    /complement(1294..3900)
                    /gene="g111962"
                    /complement(1294..3900)
                    /note="ORF ID:g111962
                    unknown protein"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="BAC89903.1"
                    /db_xref="GI:35212529"
                    /translation="MIFFGQKRRRTMIFAVASVPCIMAAQVLAIPVPIPAQOTV
                    AABGVQLLGLMKSASDLDGSLPMDIAPPAPOTHLQADPPQSPADTQSGAPTA
                    ADASGDIDGVSATVRPRTRRTATATATYDAPFRLALVTYDALVQVPGFS
                    LPALGAVRNSYLRGFDORFOLRLDGIISLQSRSSNNRVDISRLSLIEDIERVEVVTG

```

```

GATIRAGAGAGGATNLISSTPKPKPLTLTKRAGSGVGRSVYKGGGDDPTFVNLV
YTCIYVAFNNPYSATITPNTAQTFRERDAVAPNGSLRGYIKPELGPVATISGINDTGN
ASDSYSGKVFPRDPANRLTLRLNQNSKREAFGRSAPGLCRGGTSTVQPGNTSGR
RFLPLDANENELPCDTQAFI ENTPTAIFSGRVNFDANAATVAI PTGQSYLLEPLGT
TTFORFLQATEATVMDYDTITPTTSLNSYSLVRLLTTEPSPFPFDTAAGTGT
PGALVTPPAQPTIDNORQAOLANSQISPGONSFGINFLERDSYOOQSGADATF
PRAISRLPLIIDI SFNSNELKANGLRITNSSGSVLTTPAAGRSVPPNVLISRA
NYSYFNAPALSNLPTATGSNIILANPGLRPEGSTTYDVGDLTPARNLGIRPTFENYL
DGAIGSVSLANPNTATPAAI PLVLOQNDITRAGSGLI EAVADRPTSQLRLQIWTNS
DVRAAGNANPQSOSTPYTYQFDPGI PENSULNSASVYHQCWLVALVGRYDGGKRA
TTRIGRLELGS"
/gene="g111963"
/complement(3934. .6291)
/gene="g111963"
/complement(3934. .6291)
/notes="ORF ID:g111963
unknown protein"
/codon start=1
/transl_table=11
/protein_id="BAC89904.1"
/db xref="GI:35212530"
/translacion="MKRRNVLGISLILLVGGAGKLAAGATKIAKPYIOPGDAGDFG
TADQVIAMOTRANGBDLYTVDGKPTPRYGRVIAGRMVNHLISDPLPOPSG
AAMVHALLIGLEYETRVYCVGRPGTPIEGPCASPTKRKSGASELIVGDOGFEHP
STOKRYLHKQARIYAMSRLELKEPPTPALPRDPFILATGDNVYIKGSEGRNDY
NMFVMSDVACDQAPLARSLEPLTYAIGNHDIAGADI VNLADAPVNPALPAAGE
RSGALEGDDALAEFNNYIPLNGPQGVPOYI FDDACRAGFTPAVAGAKYRSAPK
IEARASTVPTLPGPRQIDHMSNFPDYANAHFVFLDANPLFPARVGAAGSADK
SPFGSPYPSLIDMLADLRSEOLMEFVVYHOPAFSATATRDYOMALIRVEDH
GANLVNGHVHYORTPLRAGRVAEADPTGEPAVEVDTA PDIGIRVPLGLIHI
EESGFTNGKLEDSAVGRSRSEALPGCPAGNWTHTLCAKAPRYANNGRGKI
TARFKARVSPGVNVDGDLITIQISEPLATCSASADPFYGSADAGKLADLA
DPLDIPATLVELVPAITGLTALDKFTVTRPDLTGLSVELHVPPTGTSQTLPCVKR
LVNRSYALNGVQVILGPTGYTFAVRGKERCTOHERSVVTVGLKEGAKKTVIE
CVVAEAPPDVQOAITASVSTALPVAKVPVLQOQSARATAVR"
complement(6288. .7748)
/gene="g111964"
/complement(6288. .7748)
/gene="g111964"
/notes="ORF ID:g111964"
/codon start=1
/transl_table=11
/product="dimethylalanine monooxygenase"
/protein_id="BAC89905.1"
/db xref="GI:35212531"
/translacion="MKRCKAVIYGGISGIVTAKCLRDGHOVTLIYESTDQYGIWVY
RKTSGTFSVRFONSKYLSARSDYMPROMSDPHHTLILAYINSYDHFRLRECIK
LINCQVEKRSRSDHKVTYSTPEGAASESFDALATCSGVFRPRMPTIPEGADPFGTL
LHAKYKERSMPANKRQVVMGASGVDIAYPATDRQVITSFRNEMLVPRYRAGR
PLDCNLSRLYALIPAGVRKSLYSRKPAPLWEARQCNLEPAGLSISANEFYIDL
VAKGAIETPTIAGSGQVRLFTDGSSTPADIVITAYGVGSFPPFDASVAVNHGT
DIYKHYHFDLPNGCFIIGIRIYIGALLPCAEQAMRFSLEVOVLTPTSEMRCLIO
RMRADOQKDMVASGRSFOVROVEYETEARLIDAVPKVMHMQALQTLTPMLPTH
YLDGPNPKGAEWIRQVPLDIKIPRNASGVGR"
8491. .12084
/gene="g111965"
/notes="ORF ID:g111965"
/codon start=1
/transl_table=11
/product="WD-repeat protein"
/protein_id="BAC89906.1"
/db xref="GI:35212532"
/translacion="MKTPKRRKRGVLTPOGYSRLQEARRLFRGNRGNRYTIDAM
SRRTGLAPRTLSKMSRKGVDROTLQECFHTPGIOLNEDVLSIOLIGRGRQSHAK
PMHELEGANPQAFYGRBEGLATLKQMTIOERSRLYAVTGMGIGTCTICABELQOIH
LEFEYVLMRSVRNGPLPKELLADLVRFESHQELBLVAFSEQVYSLAHKRLKHLV
LVLDNETMRSGLAGRYRPGREYGLFAGVASTAHNSCLVITSCBOPLKRNKPLCV
RLPVLRLYRGGIGYHGRQIILEKGLIABEDGPWHEKELRCYAGNPLAKIYASTIRE
LEDGOANPLQOOSVAFGIGIYDLLEOHYRLDIEBCLYMLAVNREPVLISQLOEDV
LLPVSRLRLTEMESLRASLIECGGTRTTLTPVLETTYTDLALAIRLELVDSYVL

```

```

PRPARILKASDEPHVKAQIRFLKPLVDOLLAVERSRRLAEHYTHRLIGELRSELO
QUGYAVANLNLHCIGIELGSDYSTLTLYRQADIRGALCHNARQNTPSQSTFTEA
ISTVSVASPDGQULATSEINGTIRLWQAAQOULAYRGHTSWMSIAFSPDGVLA
AGSADRTLRMLDRTYGOCLKVOEGBGVRSVAFLPAGGIIASGSEDAALMEVDS
GACILTLRSHSMIHA VRSPNGOMLASSDQGXOLMHPESGEPLQANQHTGVRS
IAPAPGQTLSDQDOTRLMDVDORGLIKCLOGHTGVRSVDFPADGRTLASGSD
OTVRLMDDSGICSPYMHGHSWISSVPSPGRLITSGSVSVIWEIISGHCIRV
LGHSAGIRSVAFREDGKTLASGSDHSYRLMDFSTRQMRSLQHTSVRYTASPD
GTULASGQDRTIKMDPDSGRCLTLRGTWVNSLAFPMGALIASSVHSRLTW
NVEIQCLQMLQHTHSWASVAFHDPGRVLASQDKTARLIERGLWTLQHTS
WRSYVAFHDPDHTLASGSDGVTKLMDVQTLASDLSGSGSVMSVAAAGKRLAS
GGDDKTVRLMDTSMQCTHVRHSGVLCVIAIEADSRILASSADETITLMDLQGN
YGTMRIBGPTVNMITGATGISEAQATLIXALGAVERNAAAPATPSGIDR"
complement(12067. .12756)
/gene="g111966"
/complement(12067. .12756)
/notes="ORF ID:g111966
hypoetical protein"
/codon start=1
/transl_table=11
/protein_id="BAC89907.1"
/db xref="GI:35212533"
/translacion="MGDRGLAVIALVTLVAGLSPALGSLPATAPLAVAGLSFF
IDQFLQGRSLVLYGLIOERBGRREVAAREACHILVAHRLGLPVBGYTGAMESF
RFGQGGGVIIITGSPPARIGETIGSYCATLALGALAEHMYTSEAVGAVEDQKGL
LAAORTGVAAARTLRNRAERRAAALIDBERQTHNALARMULTGOSLERCIOLNSTK
DAASDSAAVA"
complement(12761. .13045)
/gene="g111967"
/complement(12761. .13045)
/notes="ORF ID:g111967
hypoetical protein"
/codon start=1
/transl_table=11
/protein_id="BAC89908.1"
/db xref="GI:35212534"
/translacion="MPKATWNGTVALAQSDKCEVVEGNVYFPPDIAIOTEXFQPSNRTTT
CPMKGVASYDVVDVDSKVNKODAMWYFPFKDAKHITGVIAWKGRVBA"
complement(13242. .14666)
/gene="g111968"
/complement(13242. .14666)
/gene="g111968"
/notes="ORF ID:g111968
unknown protein"
/codon start=1
/transl_table=11
/protein_id="BAC89909.1"

```

Query Match 7.5%; Score 104.8; DB 1; Length 299750;
 Best Local Similarity 52.2%; Pred. No. 4,9e-10;
 Matches 260; Conservative 0; Mismatches 232; Indels 6; Gaps 1;

QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB		
624	CCAAATCGCGGAGACGCTTTTGTTCGGTGAACCCCGCGCGCAATTCGCTCAACGCT	684	GGAAGCCGGGACCACTTTTTCGTTATTCGTCGGCGAGCCGTGGAACGAGCGCGTGT	744	CAAGGCGATAGGCGTGCCTGGTATGATCAACGATGAAGAATCTTGAAAGACACCGCGGAAA	804	AATCTTCGCGCGGACCAACATATGGCGGAGAAATACCCCAACACTATCTTGGCGGTGC	864	CAAGGCGCTGATTCGGGCGCGCATCTGGCTGACCGCGCAATTAACAAGAACCGAAGGA	923	142707	CAAGCGCTCATCAAGC-----CTGCTTGTACTGCGCGCGGAGGAGGAGCGCGGAGGA	142760
142467	CAACCGGACCGGACGREGATCTACACGCTGCGCGCGCCAGATGATTGCCAACC	142527	CAAGGCGGACCACTTCGACGCTTTTTCATTCGGGCAACCTTGAAACATCCGGGCGCGC	142587	CGAGGCGATTCGCTTTTACCGGTGCGCAACGATCTTGAGAGATCTGCGCGGTCACCCGGGCA	142646	142647	GGTGCTGGGGGCGCGCGAGCATGGGCGGACCTAACCCCAATACCCACCTGGCCCTCAT	142706	142707	CAAGCGCTCATCAAGC-----CTGCTTGTACTGCGCGCGGAGGAGGAGCGCGGAGGA	142760	

```
QY      924 AGCCATCGAAATGCTGGCGCAAAAACATACGTGCGTGTGACGTGGAAGTGTGGCGGC 983
Db      142761 GATCCCGGCGCTTCTTTGCCGACGAGTACTGGGGGCCGATCCGCTACGTCCAGTT 142820
QY      984 TAGCATGAACGGCACTTTGATATCGAAAAAGACGATAACGGCGGCTACCGACTTCAA 1043
Db      142821 GGGCGAGGGCGAAACCGAACACTGTGCTCTGGCAAGACCGCTCCAGTACGCCACCA 142880
QY      1044 CACCTTCTTTGCGCACGGCGCCAGCTATCCGTCCTACAGCAGTGCAGTGTATCTGAC 1103
Db      142881 CCGTTTGGCGGTGCGCGGTGAACCGCCCGACGCGCACCGAGCACTGTGATCTCAC 142940
QY      1104 CCACTGAGGCGCTGGGG 1121
Db      142941 CCAGATGGCTCGCTGGGG 142958
```

Search completed: May 20, 2005, 21:55:48
Job time : 6184 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2005, 05:34:38 ; Search time 811 Seconds
(without alignments)
10160.641 Million cell updates/sec

Title: US-10-689-200-1

Perfect score: 1392
Sequence: 1 atgaaccatcatgagtc.....ccgacgacgaagctctgcat 1392

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16dec04:*
1: geneseqn19808:*
2: geneseqn19908:*
3: geneseqn20008:*
4: geneseqn20018:*
5: geneseqn20028:*
6: geneseqn20038:*
7: geneseqn20048:*
8: geneseqn20058:*
9: geneseqn20068:*
10: geneseqn20078:*
11: geneseqn20088:*
12: geneseqn20098:*
13: geneseqn20108:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1392	100.0	1392	12	ADO06542 Methylo
2	143	10.3	551	14	ADO06544 Methylo
3	85.4	6.1	747	11	ABD14925 Pseudom
4	85.4	6.1	750	11	ABD15326 Pseudom
5	85.4	6.1	831	11	ABD15218 Pseudom
6	82.4	5.9	1371	11	ACH97397 Kiebiell
7	71.8	5.2	1209	8	ACA45921 Pseudom
8	69	5.0	594	11	ABD15217 Pseudom
9	69	5.0	3720	11	ABD15023 Pseudom
10	62.2	4.5	1173	8	ACA24128 Prokaryot
11	59.2	4.3	576	8	ACA23622 Prokaryot
12	58	4.2	1200	8	ACA42316 Prokaryot
13	58	4.2	1200	13	ADSI14594 Pseudom
14	58	4.2	1260	11	ABD10532 Pseudom
15	57	4.1	1116	8	ACA44309 Prokaryot
16	50	3.6	1011	9	ADA32159 DNA encod
17	49.4	3.5	1512	13	ADP43829 Bacteri
18	48.4	3.5	773	8	ACA20473 Prokaryot
19	48	3.4	744	13	ADT44805 Bacteri
20	47.8	3.4	2000	8	ADA71938 Rice gene

21	47.8	3.4	7656	8	ACA37618 Prokaryot
22	47.8	3.4	28804	2	AAT37329 Sphingom
23	47.8	3.4	28804	2	AAT92474 Sphingom
24	47.8	3.4	28804	2	AAV99812 Sphingom
25	47.8	3.4	28804	2	AAV9812 Chromosom
26	46.2	3.3	1044	12	ADJ35031 DNA encod
27	46.2	3.3	37507	12	ADH48030 Clon
28	45.6	3.3	1173	8	ACA25903 Prokaryot
29	45	3.2	8298	4	AAK72613 Human
30	44.8	3.2	1333	12	ADM99142 Environm
31	43.4	3.1	420	11	ABD15254 Pseudom
32	43.4	3.1	531	11	ABD15367 Pseudom
33	43.4	3.1	1199	11	ABD15153 Pseudom
34	43.4	3.1	1188	8	ACA25753 Prokaryot
35	43.4	3.1	1389	13	ABD14772 Pseudom
36	43.4	3.1	5577	13	ADR44774 Bacteri
37	43	3.1	1305	8	ACA23608 Prokaryot
38	43	3.1	1308	11	ABD03949 Pseudom
39	43	3.1	1359	11	ABD03676 Pseudom
40	43	3.1	1443	11	ABD04027 Pseudom
41	43	3.1	1818	13	ADS48801 Bacteri
42	42.8	3.1	516	11	ABD10614 Pseudom
43	42.8	3.1	651	11	ACH94942 Kiebiell
44	42.6	3.1	1189	11	ABD03368 Pseudom
45	42.6	3.1	1587	11	ABD03299 Pseudom

ALIGNMENTS

RESULT 1	ADO06542	ADO06542 standard; DNA; 1392 BP.
ID	ADO06542	
XX	ADO06542	
AC	ADO06542	
XX	ADO06542	
DT	29-JUL-2004	(first entry)
XX		
DE	Methylomonas sp 16a nitrogen transporter ntrA gene.	
XX		
KW	de; gene; C1 metabolizing bacteria; gene expression regulation;	
KW	nitrogen transporter; ntrA.	
XX		
OS	Methylomonas sp.; 16a.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1392
FT		/*tag= a
FT		/product= "nitrogen transporter ntrA"
FT		/partial
FT		/note= "no stop codon"
XX		
PN	WO2004037998-A2.	
XX		
PD	06-MAY-2004.	
XX		
XX	21-OCT-2003; 2003WO-US033698.	
PF		
XX	21-OCT-2002; 2002US-0419872P.	
PR		
XX		
XX	(DUPO) DU PONT DE NEMOURS & CO E I.	
PA		
XX	Dicoimo DJ, Picataggio SK, Seip JE, Ye RW, Wang T, Ni H;	
PI		
XX	WPI, 2004-375911/35.	
DR	P-PDB; ADO06543.	
XX		
PT	Promoter regions responsive to presence of nitrate, acidic pH, elevated	
PT	temperatures or highly expressed in presence of methane or methanol	
PT	useful for driving expression (especially for producing zeaxanthin) in C1	
PT	metabolizing bacteria.	
XX		
PS	Claim 16; Page 69-70; 83pp; English.	

XX The present invention relates to a method of expressing a coding region
 CC of interest in a CI metabolizing bacteria. Also provided are protein,
 CC coding and promoter sequences from *Methylobacterium* sp. 16a which respond to
 CC various metabolic and growth conditions, which can be used in the method
 CC of the invention. The promoters are useful for driving expression of a
 CC number of coding regions (especially for producing zeaxanthin) in CI
 CC metabolizing bacteria under controlled conditions. The present sequence
 CC is a *Methylobacterium* sp. coding sequence.

XX Sequence 1392 BP; 347 A; 381 C; 405 G; 259 T; 0 U; 0 Other;

XX Query Match 100.0%; Score 1392; DB 12; Length 1392;

XX Best Local Similarity 100.0%; Pred.No.0;

XX Matches 1392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAACCATCTTATGATCGAGCTGAGAAATTTGTTATGAGTTATCGCTTCTGCTA 60
 DB 1 ATGAAAACCATCTTATGATCGAGCTGAGAAATTTGTTATGAGTTATCGCTTCTGCTA 60
 QY 61 GCGGTTGGGGTTTGAAGATTCGCCGATGTCGGGCGAGTTGGCAAGCTGGAAAAGAA 120
 DB 61 GCGGTTGGGGTTTGAAGATTCGCCGATGTCGGGCGAGTTGGCAAGCTGGAAAAGAA 120
 QY 121 GATTTGAAATTCGGCTTCAATCAAGCTCACCGACATGGCGCGCTGGCGGTGGCGCGAA 180
 DB 121 GATTTGAAATTCGGCTTCAATCAAGCTCACCGACATGGCGCGCTGGCGGTGGCGCGAA 180
 QY 181 AAAGCTTCTTCAAGAGACGAGGCGCTGTTCTGTCGAATGGAAGCGAGCGCAACTGGAAG 240
 DB 181 AAAGCTTCTTCAAGAGACGAGGCGCTGTTCTGTCGAATGGAAGCGAGCGCAACTGGAAG 240
 QY 241 GTGGTATGATGATGGGTCTGTAATGGCGAACTGGAAGCGCTCGCAATGCTGCGCGCGG 300
 DB 241 GTGGTATGATGATGGGTCTGTAATGGCGAACTGGAAGCGCTCGCAATGCTGCGCGCGG 300
 QY 301 CGGTTAGCGGCGAGGCTTGGCTTCGCGACAAAGGCGGATATGAGGTTGCGCTTACAGATG 360
 DB 301 CGGTTAGCGGCGAGGCTTGGCTTCGCGACAAAGGCGGATATGAGGTTGCGCTTACAGATG 360
 QY 361 GCGCTTCAACGGGACGCGATTAAGGTGTCCAAATGAATCTGGATCAGATGAAGCCGAAC 420
 DB 361 GCGCTTCAACGGGACGCGATTAAGGTGTCCAAATGAATCTGGATCAGATGAAGCCGAAC 420
 QY 421 ATACCGCTGGAAGGCGGTAAACCGGTGATTCGATCAAGCGAGATTAATCTAAACCGGT 480
 DB 421 ATACCGCTGGAAGGCGGTAAACCGGTGATTCGATCAAGCGAGATTAATCTAAACCGGT 480
 QY 481 GTGAAAATTAACAAAGCCGAGGCGGTTCAATATGCGCATGACCTTCCCGGCGGA 540
 DB 481 GTGAAAATTAACAAAGCCGAGGCGGTTCAATATGCGCATGACCTTCCCGGCGGA 540
 QY 541 TCTCAACATCAAACTGCGTTACTGCTGCGGCTGCGGCTGATCAATCTTGGCTACTAT 600
 DB 541 TCTCAACATCAAACTGCGTTACTGCTGCGGCTGCGGCTGATCAATCTTGGCTACTAT 600
 QY 601 TGGCGGCGCGAGGACATTTCCGCGCAATTCGCGCGAGAGCGCTTGTTCGTGACCCG 660
 DB 601 TGGCGGCGCGAGGACATTTCCGCGCAATTCGCGCGAGAGCGCTTGTTCGTGACCCG 660
 QY 661 CGGCGGCAAAATGCGGTCAAGCTGGAAGGCGGACCAATTTTGGTTATTTGGCGCGAG 720
 DB 661 CGGCGGCAAAATGCGGTCAAGCTGGAAGGCGGACCAATTTTGGTTATTTGGCGCGAG 720
 QY 721 CCGTGAACCAAGGCGGTGTTCAAGGCGATAGCGTCCGGTGATCAACCGTGAAGAA 780
 DB 721 CCGTGAACCAAGGCGGTGTTCAAGGCGATAGCGTCCGGTGATCAACCGTGAAGAA 780
 QY 781 CTCTGGAAGGACACGCGGGAAGAAAGTCTTCGCGGTGACCAAAATGCGCGGGAATATAC 840
 DB 781 CTCTGGAAGGACACGCGGGAAGAAAGTCTTCGCGGTGACCAAAATGCGCGGGAATATAC 840
 QY 841 CCCAACACTATCTGCGGTGACCAAGGCGGTGATTCGCGCGCGCATCTGCGTGAAGCC 900

DB 841 CCCAACACTATCTGCGGTGACCAAGGCGGTGATTCGCGCGCGCATCTGCGTGAAGCC 900
 QY 901 GACATTAACAGAACCCGAGAGAGCCATCGAATCTGCGCGCAAAAACATACGTGCT 960
 DB 901 GACATTAACAGAACCCGAGAGAGCCATCGAATCTGCGCGCAAAAACATACGTGCT 960
 QY 961 GCTGACGTGAAGTCTGCGCGGTGACATGAACCGGCACTTTGGAATTCGAAAAGACGAT 1020
 DB 961 GCTGACGTGAAGTCTGCGCGGTGACATGAACCGGCACTTTGGAATTCGAAAAGACGAT 1020
 QY 1021 AAACGCGGCTACCGGACTTCAACACCTTCTTTCGCGACGCGCGAGCTATCGCTCTAC 1080
 DB 1021 AAACGCGGCTACCGGACTTCAACACCTTCTTTCGCGACGCGCGAGCTATCGCTCTAC 1080
 QY 1081 AGCAGTGCAGTCTGATCTGACCCAGCTGAGGCGCTGGGCGATGATCAATGAA 1140
 DB 1081 AGCAGTGCAGTCTGATCTGACCCAGCTGAGGCGCTGGGCGATGATCAATGAA 1140
 QY 1141 CCGGACACTGCTATCTGATACCGCGCAAGACGTTACCGCGCGACATCTATCTGCGC 1200
 DB 1141 CCGGACACTGCTATCTGATACCGCGCAAGACGTTACCGCGCGACATCTATCTGCGC 1200
 QY 1201 GCGGCGCAAGAACTGGTGCAGAGAGCAAGGCGCAAGGCGGAACTTCCCTGCGCATCC 1260
 DB 1201 GCGGCGCAAGAACTGGTGCAGAGAGCAAGGCGGCAAGGCGGAACTTCCCTGCGCATCC 1260
 QY 1261 TCGATCAACCGCTCGCAGAAATTTCTTCATCGACAAAGTCCGTTGATCCCAAGAGCC 1320
 DB 1261 TCGATCAACCGCTCGCAGAAATTTCTTCATCGACAAAGTCCGTTGATCCCAAGAGCC 1320
 QY 1321 AACGATTACTCTGCGCAAGTTTTCGATAGGTCTGAAAGGCAAGCAACCGTAGCCGCGG 1380
 DB 1321 AACGATTACTCTGCGCAAGTTTTCGATAGGTCTGAAAGGCAAGCAACCGTAGCCGCGG 1380
 QY 1381 AAGGTGCTGAT 1392
 DB 1381 AAGGTGCTGAT 1392

RESULT 2
 ADO06544
 ID ADO06544 standard; DNA; 551 BP.
 XX ADO06544;
 AC
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE *Methylobacterium* sp 16a nitrogen transporter nrtA gene promoter.
 XX
 KW ds; gene; CI metabolizing bacteria; gene expression regulation;
 XX nitrogen transporter; nrtA; promoter.
 XX
 OS *Methylobacterium* sp.; 16a.
 XX
 PN WO2004037998-A2.
 XX
 PD 06-MAY-2004.
 XX
 PF 21-OCT-2003; 2003WO-US033698.
 XX
 PR 21-OCT-2002; 2002US-0419872P.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Dicosimo DJ, Picataggio SK, Seip JE, Ye RW, Wang T, Ni H;
 XX WPI; 2004-375911/35.
 XX
 PT Promoter regions responsive to presence of nitrate, acidic pH, elevated
 PT temperatures or highly expressed in presence of methane or methanol
 PT useful for driving expression (especially for producing zeaxanthin) in CI
 PT metabolizing bacteria.

XX Claim 37, Page 72, 83pp; English.
PS
XX The present invention relates to a method of expressing a coding region
CC of interest in a CI metabolizing bacteria. Also provided are protein,
CC coding and promoter sequences from *Methylobacterium* sp. 16a which respond to
CC various metabolic and growth conditions, which can be used in the method
CC of the invention. The promoters are useful for driving expression of a
CC number of coding regions (especially for producing zeaxanthin) in CI
CC metabolizing bacteria under controlled conditions. The present sequence
CC is a *Methylobacterium* sp. promoter sequence.
SQ Sequence 551 BP; 139 A; 119 C; 135 G; 158 T; 0 U; 0 Other;
Query Match 10.3%; Score 143; DB 12; Length 551;
Best Local Similarity 100.0%; Pred. No. 7e-27;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAACCTCATTTAGATGAGCTCGAAGAAATTGTTATGACGTTATCGGCTTGGCTA 60
DB 409 ATGAAACCTCATTTAGATGAGCTCGAAGAAATTGTTATGACGTTATCGGCTTGGCTA 468
QY 61 GCGGTTGGGCTTGAAGATTGGCCCGAGTGGGGGAGTTGGCAAGCTGAAAGAA 120
DB 469 GCGGTTGGGCTTGAAGATTGGCCCGAGTGGGGGAGTTGGCAAGCTGAAAGAA 528
QY 121 GATTGAATTCGGCTTCATCAA 143
DB 529 GATTGAATTCGGCTTCATCAA 551
RESULT 3
ABD14925/c
ID ABD14925 standard; DNA: 747 BP.
AC ABD14925;
XX
DT 29-JUL-2004 (first entry)
XX
DE *Pseudomonas aeruginosa* polynucleotide #13529.
XX
KW Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;
XX antibacterial.
XX
OS *Pseudomonas aeruginosa*.
XX
PN US651795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
XX
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR P-PSDB; ABO81354.
XX
XX Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 13529; 455bp; English.
XX
XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,

CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-*P. aeruginosa* drugs, as templates for recombinant
CC production of *P. aeruginosa*-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of *P. aeruginosa*-caused
CC infection, and in detection of *P. aeruginosa* sequences or other sequences
CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-
CC ABD1967 represent *P. aeruginosa* polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 747 BP; 114 A; 234 C; 281 G; 118 T; 0 U; 0 Other;
Query Match 6.1%; Score 85.4; DB 11; Length 747;
Best Local Similarity 51.3%; Pred. No. 7.5e-12;
Matches 250; Conservative 0; Mismatches 231; Indels 6; Gaps 2;
QY 651 GGTGACCCCGCCCGCCCAATGCGCTTCACGCTGGAAGCCGCACTTTCCGTTATTG 710
DB 559 GGTGATTCACCCCGCCAGATGCTGCGCATCTCCAGGCGCGGCGCATCGAGGTTCTG 500
QY 711 CGTGGCGAGCGGTGGAACGAGAGCCGTTCAAGGCGATAGGCGTGGCGGATCAC 770
DB 499 CGCGGGGCGCCCTGGGGCGCCCTGGCGGTGACAGGCGAGGCTTCACTTCGCGCAC 440
QY 771 CGATGAAGAACTGGAAGAGCACGCGGAAAAAGTCTTGGCGGTGACCAACATAGGCG 830
DB 439 CAGCCAGGCGATTTGGCCGAGCACCCGGAAGAGTCTTGGGACATACCGGCGCTTCGT 380
QY 831 GGAAGAAATCCCAACACCTATCTGGCGGTGACCAAGCGGTGATTCGGCGCGGATCTG 890
DB 379 CGAGGCTTACCCGAAACACGCGCGCGCTGTGATGAGC--GATCTCGAGCGCCAGCCG 323
QY 891 GCTGAGCGCCGGAATTAACAGAACCGCAAGAGCATGAAATGCTGGCGGAAAAACA 950
DB 322 CTTCATGGAACAGAAACGCGAGAACCGCTGGGACCGGCAACTGATCAGCGCGCGCA 263
QY 951 ATAGCTGCGTGTGAGTGAAGTGTGCGGCGCTAGCATGAAGCGCACTTTCGAATACGA 1010
DB 262 CTATGTGAGCTTCGCTGCGCGCATCCAGCGCGCTTCTTGGCGCTACAGAGACGG 203
QY 1011 AAAAGACGATTAACGCGCGCTACCGGACCTTCAACACTTCTTTCGCCAGCGGCCAGCTA 1070
DB 202 CTCGGCAACGCTGGGAGAGACCAATCCGCTGCGTTCTTACGCGGCAAGTCAA 143
QY 1071 TCCGTCT--AAGCAGTCACTGTGATCTGATCCAGCTAGAGCGCTGGGCGATGAT 1127
DB 142 CGGCGCTGCGCTCTCGAATGCGATGTGTTCAATGACCACTTCGCGCTGGGCGCTGCT 83
QY 1128 CAATGAA 1134
DB 82 GCGGAA 76
RESULT 4
ABD15326
ID ABD15326 standard; DNA: 750 BP.
AC ABD15326;
XX
DT 29-JUL-2004 (first entry)
XX
DE *Pseudomonas aeruginosa* polynucleotide #13930.
XX
KW Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;
XX antibacterial.
XX
OS *Pseudomonas aeruginosa*.
XX
PN US651795-B1.
XX
PD 22-APR-2003.

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-regulated gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ. Sequence 1209 BP; 229 A; 395 C; 376 G; 209 T; 0 U; 0 Other;

Query Match 5.2%; Score 71.8; DB 8; Length 1209;

Best Local Similarity 53.8%; Pred. No. 3e-08; Matches 188; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 102 TGGCAAGCTGGAAAAGAAAGATTGAATTCGGCTTCATCACTGACCGACATGGCGGC 161
DB 45 TGAACGCCCGCGAATGAACAGCTGGACGTGGGCTTCATGACGACGACGACGCGCC 104
QY 162 GCTGGGGGTGGCCCGCGAAAAGCTTCTTCGAGACGAGGCGCTTCTGTCGACATCGA 221
DB 105 GCTGGGTGGTGGTGGCCACCCAGGGGTTTGGCCGCTTACGGGCTGAGCCTGAACCTGA 164
QY 222 AGCGCAGGCAATCTGAAAGGTGTGATGTAGTGGTCTGTATGATGAGGAACTGAGCGGCTC 281
DB 165 AGCCGAGACTCTCTGGGCGCGGCTGCTGACAGAGCTGTACGGGCGAGCTTAAAGCGGC 224
QY 282 GCACATGCTGGCGCCGCGCGCTTACCGGCGACGCTTGGCTTGGGACCAAGGCCGATAT 341
DB 225 ACATAGCTGTATGTTGATGTATGCGGAGAGCTGGGATCAGCGGCGCGCGAC 284
QY 342 CGAGGTGCGCTTACGATGGGCTTCAACGGCAACG 376
DB 285 CGACATGGCCATCTCATGGGCTTCAACGAACG 319

RESULT 8

ABD15217

ID ABD15217 standard; DNA; 594 BP.

XX ABD15217;

DT 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polymucleotide #13821.

XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;

KW antibacterial.

XX Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX

PF 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

PI Rubenfield MJ, Nolling J, DeLoughery C, Bush D;

XX MPI; 2003-615309/58.

DR P-PSDB; ABO81646.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,

PT useful as molecular targets for diagnostics, prophylaxis and treatment of

PS pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 13821; 455bp; English.

CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biophysics technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html

XX SQ. Sequence 594 BP; 102 A; 208 C; 196 G; 88 T; 0 U; 0 Other;

Query Match 5.0%; Score 69; DB 11; Length 594;

Best Local Similarity 52.2%; Pred. No. 1.3e-07; Matches 153; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 84 GCCCGATGTCGGGGGAGTGGCAAGCTGGAAAAGAAAGATTGAATTCGGCTTCATCA 143
DB 204 GGGCTGGGCGCGCGGCTTCGAGACGACCGAAAAGAGCCCTGGACATCGTTTCAATGCG 263
QY 144 GCTCACCGCATGAGCGCGCTGAGCGTGGCGCCGCAAAAAGCTTCTTTCGAGACGAGG 203
DB 264 GCTGACCGATGTCGGCTGCTGATGCTGAGCGCCGACCCAGGCTTGGCCAGCCCTAAG 323
QY 204 CCTGTTTCGTCACATGGAAGCGGCACTGGAAGGTGTGATGATGATGAGGTGCTGA 263
DB 324 CCTCAGGCTCAATCTTCAAGCGGCAACCGTCTGGGCGACCTTGGCGGACAACTGCTCA 383
QY 264 TGGCGAATGAGCGGCTGCGCATGCTGGCGCGCGCGCTTACGCGGCGAGGCTTGGCTT 323
DB 384 CGCGAGCTGATGATGCGCGGCAAGTGTGATGAGGCTTGTCTACGAGCTCAAGCTCGGCT 443
QY 324 CGGACCAAGCGCGATATGAGGTGCGCTTCAAGATGGGCTTCAACGGGCAACG 376
DB 444 CGGCGGCGAGCGGCGGCAAGGATGGCGGTGCTCATGAGACTGTGCAAAAAG 496

RESULT 9

ABD15023/c

ID ABD15023 standard; DNA; 3720 BP.

XX ABD15023;

DT 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polymucleotide #13627.

XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;

KW

XX antimicrobial.
XX Pseudomonas aeruginosa.
OS US6551795-B1.
PN 22-APR-2003.
XX 18-FEB-1999; 99US-00252991.
XX 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
PI WPI: 2003-615309/58.
XX P-PSDB; AB081452.
DR Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
XX Disclosure, SEQ ID NO 13627; 455bp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 3720 BP; 576 A; 1147 C; 1326 G; 671 T; 0 U; 0 Other;
Query Match 5.0%; Score 69; DB 11; Length 3720;
Best Local Similarity 52.2%; Pred. No. 2.4e-07;
Matches 153; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 84 GCCCGATGTCGGG3CAAGTTGGCAAGTGAAGAAAGAAAGATTGCGGCTTCATCAA 143
DB 676 GCGCTGGGCGCGGGGTTCCGACCGACGGAAGAGGCGCTTGACATCGGTTTCATGCGC 617
QY 144 GCTCACCGACATGCGCGCGCTGGCGGTGCGCGCGGCAAAAGGCTTCTTCAGACGAGGG 203
DB 616 GCTGACCGAATCGGCGCTCGCTGATCGGCGCGCACACCAAGGGGCTTCCGCCAGCCTTACGG 557
QY 204 CCGTTCGTGTCACATGGAAGCGACGCGCAACTGGAAGTGTGATGATGATGAGTGGTGGAA 263
DB 556 CCTCACGCTCAATCTCAGGCGCCAAACGTCCTGGGCCACCTTGGCCACAAGCTGCTCAG 497
QY 264 TGGCGAATGAGAGGCTGCGCATGCTGCGCGCGCGCGCTTGAAGCGGCCAGCGTTGGCTT 323
DB 496 CCGCGAGCTGGAATGCGCGCGACGATGCTGTACGCGCTTGTCTACGCGCTCCAGCTCGGCTT 437
QY 324 CCGGACCAAGAGCGCGATTCAGAGGTGCGCTTCAAGATGCGGCTTCAAGCGGCAAGC 376
DB 436 CCGCGGCGAGCGCGCGCAAGGATGCGCGGTCTCATGGAGCATGTGCGCAAAACG 384

RESULT 10
ACA24128
ID ACA24128 standard; DNA; 1173 BP.

XX ACA24128;
AC 19-JUN-2003 (first entry)
XX
DT Prokaryotic essential gene #5785.
XX
DE Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX Borrelia cepacia.
XX W0200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
XX WPI: 2003-029926/02.
XX P-PSDB; AB020258.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 11998; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation; (7) identifying a gene in an operon required for
CC proliferation; (8) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway;
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism's activity; (9) manufacturing an antibiotic; (10) profiling a
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1173 BP; 180 A; 396 C; 404 G; 193 T; 0 U; 0 Other;
Query Match 4.5%; Score 62.2; DB 8; Length 1173;

Best Local Similarity 58.4%; Pred. No. 9.4e-06;
Matches 128; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

```

QY 131 TCGGCTTCATCAAGCTCACCAGACATGCGCGCTGCGCGCCGCAAAAAGCTTCT 190
    |||||
DB 116 TCGGCTATCTGCGCATCACCGACGCGGCCCGCTGCTGGTGGCCCAACAACAGGCTATT 175
    |||||
QY 191 TCGAGACGAGGGCCCTGT---TCGTCAACTGGAAGCGCAGGCCAATGTGAAGTGTGA 247
    |||||
DB 176 TCGCGCTCCAGAGGGCTGACGCTGACCAACCGAAGCTGCTGCTGCGGCCCACTCG 235
    |||||
QY 248 TGGATATGAGTTCGTAATGGAATGGAATGCACTGCAATGTGCGCGCCGCTTAA 307
    |||||
DB 236 TCGAGGCGTTCTCTGTCCGCGCAGGTCAAGCTGTGCACTGCTGTGGCCGATACGCTGT 295
    |||||
QY 308 CGGCCAGCGTTGGCTTTCGCGACCAAGCCGATATCGAG 346
    |||||
DB 296 GGGCGCGCTACGCGACCGCGCGCGCGCAAGGTGTGG 334
    |||||

```

RESULT 11

ACA23622
ID ACA23622 standard; DNA; 576 BP.

ACA23622;

19-JUN-2003 (first entry)

Prokaryotic essential gene #5279.

Antisense; ds; prokaryotic essential gene; cell proliferation;
drug design; gene.

Borrelia cepacia.

W020027183-A2.

03-OCT-2002.

21-MAR-2002; 2002MO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342823P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

WPI; 2003-029926/02.

P-PSDB; ABU19752.

New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 11492; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway

required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene
product is overexpressed or underexpressed; (12) determining the extent
to which each of the strains is present in a culture or collection of
strains; or (13) identifying the target of a compound that inhibits the
proliferation of an organism. The antisense nucleic acids are useful for
identifying proteins or screening for homologous nucleic acids required
for cellular proliferation to isolate candidate molecules for rational
drug discovery programs, or for screening homologous nucleic acids
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Sequence 576 BP; 84 A; 206 C; 194 G; 92 T; 0 U; 0 Other;

Query Match 4.3%; Score 59.2; DB 8; Length 576;

Best Local Similarity 57.9%; Pred. No. 4.5e-05;
Matches 125; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

```

QY 131 TCGGCTTCATCAAGCTCACCAGACATGCGCGCTGCGCGCCGCAAAAAGCTTCT 190
    |||||
DB 116 TCGGCTATCTGCGCATCACCGACGCGGCCCGCTGCTGGTGGCCCAACAACAGGCTATT 175
    |||||
QY 191 TCGAGACGAGGGCCCTGT---TCGTCAACTGGAAGCGCAGGCCAATGTGAAGTGTGA 247
    |||||
DB 176 TCGCGCTCCAGAGGGCTGACGCTGACCAACCGAAGCTGCTGCTGCGGCCCACTCG 235
    |||||
QY 248 TGGATATGAGTTCGTAATGGAATGGAATGCACTGCAATGTGCGCGCCGCTTAA 307
    |||||
DB 236 TCGAGGCGTTCTCTGTCCGCGCAGGTCAAGCTGTGCACTGCTGTGGCCGATACGCTGT 295
    |||||
QY 308 CGGCCAGCGTTGGCTTTCGCGACCAAGCCGATATCGAG 343
    |||||
DB 296 GGGCGCGCTACGCGACCGCGCGCGCGCGCAAGGTGTGG 331
    |||||

```

RESULT 12

ACA42316
ID ACA42316 standard; DNA; 1200 BP.

ACA42316;

19-JUN-2003 (first entry)

Prokaryotic essential gene #23973.

Antisense; ds; prokaryotic essential gene; cell proliferation;
drug design; gene.

Pseudomonas aeruginosa.

W020027183-A2.

03-OCT-2002.

21-MAR-2002; 2002MO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykind UW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR P-PSDB; ABU38446.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 30186; 1766bp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1200 BP; 190 A; 420 C; 409 G; 181 T; 0 U; 0 Other;
Query Match 4.2%; Score 58; DB 8; Length 1200;
Best Local Similarity 56.2%; Pred. No. 0.00012;
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;
QY 131 TCGGCTTCATCAGCTCAGCAGCATGCGCGCGGCGGCGCGGCGGCGGCGGCGGCGGCTTCT 190
DB 137 TCGGCTTACTCGGATATCCAGACGCCAGCCTTCTGCTGAGGCCATGCCAAGCGCTGT 196
QY 191 TCGAGACGAGGAGCGCTTCTGTCGCACTGGAAGCGCAGGCC--AACTGGAAGGTGTGTA 247
DB 197 TCGAGGCGGAGGCGGATCAGGCGGAAAGCGCGGTGCTGTTGCGTAGCTGGGCGGAGTGA 256
QY 248 TGAATAGGTCGTGAATGCGGAATGAGCGGCTCGCATGCTGCGCGCGCGCGCGCTTAG 307
DB 257 TCGAGGCGTTCATCTCCGCGCAGGTCAACGTCACTCCAGCTGTCTCCCGCATGACCTCT 316
QY 308 CGGCGAGCGTTGGCTTTCGGGACCAAGCGCATATCGAGGTCGCGTTGAGAGTGGGCTTCA 367
DB 317 GGGCC-----CGTACGGAGCAAGGTGCCGCGCAAGGTGTGGCTTGAACCAAGCTCG 370
QY 368 ACGGCAACGCGATTAACGTGTCCATGAATC 399
DB 371 GCGGCTCGGGGCTCAAGGTGCTCCGAGATC 402
RESULT 13
ADSI4594
ID ADSI4594 standard; DNA; 1200 BP.

XX AC ADSI4594;
XX 16-DEC-2004 (first entry)
XX
DE Pseudomonas aeruginosa quorum sensing controlled gene PA2328, SEQ ID 149.
XX
XX Cytostatic; Anti-HIV; Antimicrobial; Immunosuppressive;
XX quorum sensing signalling; bacterium; quorum sensing controlled gene;
XX biofilm-associated disease; cystic fibrosis; AIDS; middle ear infection;
XX acne; periodontal disease; gene; der; PA2328.
XX
OS Pseudomonas aeruginosa.
XX
XX MO2004083385-A2.
XX
XX 30-SEP-2004.
XX
XX 11-MAR-2004; 2004MO-US007467.
XX
XX 14-MAR-2003; 2003US-00389647.
XX
PA (IOWA) UNIV IOWA RES FOUND.
XX
PI Greenberg EP, Schuster M, Lostroh C;
XX WPI; 2004-709932/69.
XX
PT Identifying a modulator of quorum sensing signaling in bacteria, useful
PT for treating a biofilm-associated disorder, comprises contacting the cell
PT with a quorum sensing signal molecule in the presence and absence of a
PT test compound.
XX
PS Disclosure; SEQ ID NO 149; 233bp; English.
XX
XX The present invention relates to a method for identifying a modulator of
XX quorum sensing signalling in bacteria. The method comprises: providing a
XX cell that comprises a quorum sensing controlled gene (ADSI4446-ADSI4798),
XX where the cell is responsive to a quorum sensing signal molecule such
XX that a detectable signal is generated; contacting the cell with a quorum
XX sensing signal molecule in the presence and absence of a test compound;
XX and detecting a change in the detectable signal. The method and modulator
XX identified by the method are useful for treating a biofilm-associated
XX disease or disorder, e.g. cystic fibrosis, AIDS, middle ear infections,
XX acne, periodontal disease, catheter-associated infections, and medical
XX device-associated infections. Note: The sequence data for this patent was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1200 BP; 190 A; 420 C; 409 G; 181 T; 0 U; 0 Other;
Query Match 4.2%; Score 58; DB 13; Length 1200;
Best Local Similarity 56.2%; Pred. No. 0.00012;
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;
QY 131 TCGGCTTCATCAGCTCAGCAGCATGCGCGCGTGGCGGCGGCGGCGGCGGCGGCGGCTTCT 190
DB 137 TCGGCTTACTCGGATATCCAGACGCCAGCCTTCTGCTGAGGCCATGCCAAGCGCTGT 196
QY 191 TCGAGACGAGGAGCGCTTCTGTCGCACTGGAAGCGCAGGCC--AACTGGAAGGTGTGTA 247
DB 197 TCGAGGCGGAGGCGGATCAGGCGGAAAGCGCGGTGCTGTTGCGTAGCTGGGCGGAGTGA 256
QY 248 TGAATAGGTCGTGAATGCGGAATGAGCGGCTCGCATGCTGCGCGCGCGCGCGCTTAG 307
DB 257 TCGAGGCGTTCATCTCCGCGCAGGTCAACGTCACTCCAGCTGTCTCCCGCATGACCTCT 316
QY 308 CGGCGAGCGTTGGCTTTCGGGACCAAGCGCATATCGAGGTCGCGTTGAGAGTGGGCTTCA 367
DB 317 GGGCC-----CGTACGGAGCAAGGTGCCGCGCAAGGTGTGGCTTGAACCAAGCTCG 370
QY 368 ACGGCAACGCGATTAACGTGTCCATGAATC 399
DB 371 GCGGCTCGGGGCTCAAGGTGCTCCGAGATC 402

Dh 371 GCGGCTCGGGGCTCACGGTGGCTCCGGAGATC 402

RESULT 14
ABD10532
ID ABD10532 standard; DNA; 1260 BP.
XX
XX ABD10532;
AC
XX 29-JUL-2004 (first entry)
DT
XX Pseudomonas aeruginosa polynucleotide #9136.
DE
XX
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KM antibacterial.
XX
XX Pseudomonas aeruginosa.
OS
XX US6551795-B1.
PN
XX 22-APR-2003.
PD
XX
XX 18-FEB-1999; 99US-00252991.
PF
XX
XX 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
XX WPI; 2003-615309/58.
DR P-PSDB; ABO76961.
XX
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
XX Disclosure; SEQ ID NO 9136; 455bp; English.
PS
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant
XX production of P. aeruginosa-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of P. aeruginosa-caused
XX infection, and in detection of P. aeruginosa sequences or other sequences
XX of Pseudomonas species using biotech technology. Sequences ABD01397-
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
SQ Sequence 1260 BP; 202 A; 450 C; 423 G; 185 T; 0 U; 0 Other;

Query Match 4.2%; Score 58; DB 11; Length 1260;
Best Local Similarity 56.2%; Pred. No. 0.00012;
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

Qy 131 TCGGCTTCATCAAGCTCACCGACATGGCGCCGTCGGTGGCCGCGAAGAGCTTCT 190
Dh 197 TCGGCTACTGCGCATACCGACCGCCCTGCTGCTGGCCCATGCCACGCTCTGT 256
Qy 191 TCGAGCAGCAGGCGCTGTTCTGTCACCTGGAACCGCAGGCC--AACTGGAAGTGTGA 247
Dh 257 TCGAGGCGCAGGCGCATCCAGCGCGAAGCGCCGCTGCTGTTGCGTACCTGGCGCAGGTGA 316
Qy 248 TCGATAGGCTCGTGAATGGCGAAGTGGAGGCTCGCACTGCTGGGGCGCGCGCTTNG 307

Dh 317 TCGAGGCGTTCATCTCCGGCAGGTCAAGTCATCCACTGCTGTGCGCGATGACGCTCT 376

Qy 308 CGGCGACCGTTGGCTTCCGACCAAGCCGATATGAGGTGCCGTTACAGATGGCTTCA 367
Dh 377 GGGCC-----CGCTACGGCAGCAGAGTCCCGCAGAGGTGGGCTGTGAACCAAGTCA 430

Qy 368 ACGGCAACCGCATTTACGTGTCCATGAATC 399
Dh 431 GCGGCTCGGGGCTCACGGTGGCTCCGGAGATC 462

RESULT 15
ACA44309
ID ACA44309 standard; DNA; 1116 BP.
XX
XX ACA44309;
AC
XX
XX 19-JUN-2003 (first entry)
DT
XX
XX Prokaryotic essential gene #25966.
DE
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KM drug design; gene.
XX
XX Pseudomonas putida.
OS
XX WO2002717183-A2.
PN
XX 03-OCT-2002.
PD
XX
XX 21-MAR-2002; 2002WO-US009107.
PI
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zvekind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI
XX WPI; 2003-029926/02.
DR P-PSDB; ABU40439.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 32179; 1766bp; English.
PS
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a gene in an operon required for
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-regulated gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the

proliferation of an organism. The antitense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1116 BP; 215 A; 361 C; 342 G; 198 T; 0 U; 0 Other;

Query Match 4.1%; Score 57; DB 8; Length 1116;

Best Local Similarity 54.1%; Pred. No. 0.00021;

Matches 139; Conservative 0; Mismatches 115; Indels 3; Gaps 1;

```
QY 131 TCAGGCTTCATCAAGCTCAACGACATGCGCGCTGCGGCTGGCCGCGGAAAAAGCTTCT 190
    |||||
DB 56 TCAGGCTTCATCAAGCTCAACGACATGCGCGCTGCGGCTGGCCGCGGAAAAAGCTTCT 115
    |||||

QY 191 TCAGGAGCAGAGGCT---GTTGTCGCACTGGAAGCCGACGCACTGGAGGTGTGA 247
    |||||
DB 116 TCAGGAGCAGAGGCTCAAGGCGCGGCTGCTGCTGCGGCACTGGAGGTGTGA 175
    |||||

QY 248 TGGATAGGGTCGTGAATGGGAACTGACGCGCTCGCATGCTGCGCGCGCGCGCTTAG 307
    |||||
DB 176 TCAGGAGCATTCATCTCGGCGCAGGTCAATGTCATCCACTGCTTTCGCGCATGACGSTAT 235
    |||||

QY 308 CGGCCAGCGCTGGCTTGGGACCAAGGCGCATATCGAGTCCGTTTACGATGGGCTTCA 367
    |||||
DB 236 GGGCGGCGCTATGCGCAGCAGGTGCCAGGCCAAGGTGTGCTTGGAAACACGCTGGCGGTT 295
    |||||

QY 368 ACGGCAACGCGATTACG 384
    |||||
DB 296 CGGGCTGACTGTGGCG 312
    |||||
```

Search completed: May 20, 2005, 20:12:42
Job time : 817 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: May 20, 2005, 19:00:43 ; Search time 259 Seconds
(without alignments)
8794.198 Million cell updates/sec

Title: US-10-689-200-1

Perfect score: 1392
Sequence: 1 atgaaacacatcattagatc.....ccgcgcgaagtcgtcgat 1392

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgm2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgm2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgm2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgm2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgm2_6/prodata/1/ina/PCTUS_COMB.seq: *
6: /cgm2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	85.4	6.1	747	US-09-252-991A-13529	Sequence 13529, A
2	85.4	6.1	750	US-09-252-991A-13930	Sequence 13930, A
3	85.4	6.1	831	US-09-252-991A-13822	Sequence 13822, A
4	82.4	5.9	1371	US-09-489-039A-13192	Sequence 13192, Ap
5	69	5.0	594	US-09-252-991A-13821	Sequence 13821, A
6	69	5.0	3720	US-09-252-991A-13627	Sequence 13627, A
7	58	4.2	1260	US-09-252-991A-9136	Sequence 9136, Ap
8	50	3.6	1011	US-09-328-352-3446	Sequence 3446, Ap
9	48.2	3.5	1368	US-09-902-540-7805	Sequence 7805, Ap
C 10	48.2	3.5	6555	US-09-902-540-777	Sequence 777, App
11	48	3.4	750	US-09-902-540-5834	Sequence 5834, Ap
C 12	48	3.4	72704	US-09-902-540-1273	Sequence 1273, Ap
C 13	47.8	3.4	28804	US-08-592-874-1	Sequence 1, App11
C 14	47.8	3.4	28804	US-09-096-942-2	Sequence 2, App11
C 15	47.8	3.4	28804	US-09-096-867-2	Sequence 14, App1
C 16	44.4	3.2	7218	US-08-232-463-14	Sequence 3780, Ap
C 17	44.4	3.2	3384	US-09-902-540-3780	Sequence 1184, Ap
C 18	43.8	3.1	23233	US-09-902-540-1184	Sequence 13856, A
19	43.4	3.1	420	US-09-252-991A-13858	Sequence 13971, A
20	43.4	3.1	531	US-09-252-991A-13971	Sequence 13376, A
C 21	43.4	3.1	1119	US-09-252-991A-13757	Sequence 13376, A
C 22	43.4	3.1	1389	US-09-252-991A-13376	Sequence 2081, Ap
C 23	43.4	3.1	5523	US-09-902-540-2081	Sequence 5383, Ap
C 24	43.4	3.1	5577	US-09-902-540-5383	Sequence 1262, Ap
25	43.4	3.1	34552	US-09-902-540-1262	Sequence 2553, Ap
26	43	3.1	1308	US-09-252-991A-2553	Sequence 2280, Ap
C 27	43	3.1	1359	US-09-252-991A-2280	Sequence 2280, Ap

28	43	3.1	1443	US-09-252-991A-2631	Sequence 2631, Ap
29	42.8	3.1	516	US-09-252-991A-9218	Sequence 9218, Ap
30	42.8	3.1	651	US-09-489-039A-737	Sequence 737, App
31	42.6	3.1	1149	US-09-252-991A-1972	Sequence 1972, Ap
C 32	42.6	3.1	1587	US-09-252-991A-1903	Sequence 1903, Ap
C 33	42.6	3.1	1836	US-09-252-991A-2056	Sequence 2056, Ap
34	42.4	3.0	450	US-09-489-039A-2830	Sequence 2830, Ap
C 35	42.4	3.0	465	US-09-252-991A-1459	Sequence 1459, Ap
C 36	42.4	3.0	939	US-09-489-039A-2942	Sequence 2942, Ap
C 37	42.4	3.0	1512	US-09-252-991A-1551	Sequence 1551, Ap
C 38	42.4	3.0	1677	US-09-252-991A-1616	Sequence 1616, Ap
C 39	41.8	3.0	2025	US-09-252-991A-10293	Sequence 10293, A
40	41.8	3.0	2094	US-09-252-991A-10564	Sequence 10564, A
41	41.6	3.0	1425	US-09-540-236-972	Sequence 972, App
42	41.6	3.0	10528	US-09-902-540-945	Sequence 945, App
43	41.4	3.0	1164	US-09-902-540-5610	Sequence 5610, Ap
44	41.4	3.0	1206	US-09-902-540-4100	Sequence 4100, Ap
C 45	41.4	3.0	11812	US-09-902-540-1041	Sequence 1041, Ap

ALIGNMENTS

RESULT 1
US-09-252-991A-13529/c
; Sequence 13529, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIORITY FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIORITY FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13529
; LENGTH: 747
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-13529

Query Match	Best Local Similarity	Score	DB 4	Length	747
Matches 250; Conservative	51.3%	Pred. No. 1.3e-12;	Mismatches 231;	Indels	6; Gaps 2;
651	GGTGACCCGCGCGCGCAAAATGCGCTGCACGCTGAAGCCGACCATTTTCGTTATTG	710			
559	GGTGTCTTCCACCGCGCGCAATGCTGCGCATCTCCAGCGCGGCGATCGACGGTTCTG	500			
711	CGTGGCGAGCCGTGGAACGACGAGCGCTGTTCAAGGGATGAGCGCTGCGTATCAC	770			
499	CGCGCGCGCGCGCTGGGGCGCGCTGGCGGTGACGACGAGCGCTTACCATCGGCAC	440			
771	CGATGAAGAACTTGAAGACACGCGGAAAGTTCTTGGGTGATCAACCAATATGCG	830			
439	CAACCGAGCGATCTGGCGCGGACACCGGAAAGTCTCTGGGACATACCGCGCTTCTGT	380			
831	GGAGAAATACCCCAACCATATCTGGGGGTGACCAAGCGCTGATTCGGCGCATCTG	890			
379	CGACGCTTACCCCAACCATCGCGCGCGCGCTGCTGCTATGCG---GCTCTGACCGCGCG	323			
891	GCTGACCGCGCGCAATTAACGAAGACCGCAAGAGCATGAATGCTGCGCGCAAAACA	950			
322	CTTCATTCGAACAGAAACCGCGGACCGCTGGGACACGCGCACTATGCGCGCGCA	263			
951	ATACGTCGCGTCTGACGTGAATGCTGCGCGCTTACATGAAGGCACTTTCGAATACGA	1010			
262	CTATGTCGACGCTCGCGCTGCGCGGATTCACGCGCGCTTCTTGGCGCGTACGACGCG	203			

TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489, 039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 3192
LENGTH: 1371
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3192

Query Match 5.9%; Score 82.4; DB 4; Length 1371;
Best Local Similarity 56.2%; Pred. No. 9,7e-12;
Matches 153; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 652 GTGACCCCGCGCGCAAAATGCGTCCAGCTGGAACCGCGACCATTTGCTTATTCG 711
DB 694 GTGGTCCCGCGCGCGCGATGATGATGAACATGCGCATTCGCAATATGAGCGGCTTTGCG 753
QY 712 GTGCGGAGCGCGTGAACCAAGCGCGTGTTCAGAGGCGATGAGCGGTGATGACG 771
DB 754 GTGCGGAGCGCGTGAACCGCGCGCTATTAGACCGCTTACCGCGGCACT 813
QY 772 GATGAAGACTGTGAAGACACCGCGGAAAAGCTTTCGCGCTGACCAACATGGGCG 831
DB 814 TCCAGGATATCTGCGCGCGACATCCGAAAAGTGTGCGGACCGTGGGACTGGGTG 873
QY 832 GAGAAATACCCCAACACTTATCTGCGCGTGAACAGCGCTGATTCGGCGCGCATCTCG 891
DB 874 GAACGCAACCCCAACACCGCGCGCGCTGTGTGCGCGCGCTGATGAGGCGCGCGCTCG 933
QY 892 CTGACGCGCGCAATATACAGAACCGCAAGAGAGCC 927
DB 934 ATCGCGCTTCGCGCGGAGAACCCCGGAGAGCGCC 969

RESULT 5
US-09-252-991A-13821

Sequence 13821, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252, 991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13821
LENGTH: 594
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13821

Query Match 5.0%; Score 69; DB 4; Length 594;
Best Local Similarity 52.2%; Pred. No. 2,4e-08;
Matches 153; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 84 GCCCGATGTGGGCGCACTGGCAAGCTGGAAGGAAGATTGAATTGCGCTTCATCAA 143
DB 204 GGCGTGGCGCGCGGTTCCAGACCGGAAGAGGCGCTTGACATCGTTTCATGCG 263
QY 144 GCTACCGACATGGCGCGGCTGGCGCGCGCGGCGGAAAAGGCTTCTTCGAGACGAGGG 203
DB 264 GCTGACCGAATTCGCGCTGCTGATGCTGGCGCGGCGGCGGAGGCTTCGCCGCTTACGG 323
QY 204 CTTGTTCTGCAACTGGAAGCGGCAACTGGAAGGTGTGATGATGAGGTGCTGAA 263

DB 324 CTTACGCTAACTTCAGGCGCGCAACCGCTCTGCGCCACCTTGGCGCAAGGCTGCG 383
QY 264 TGGCGAAGTGAAGCGCTGCGACATGCTGGCGCGCGCGCTTACCGCGCAAGCTTGGCTT 323
DB 384 CGGCGAGCTGAATGCCCGGAGTGCCTGTACGGGCTTGTACGGGCTTCAGGCTCGGCT 443
QY 324 CGGCAACCAAGCGCGATATGAGGTGCGCTTACGATGAGGCTTCAACGCGCAAG 376
DB 444 CGGCGGAGCGCGCGCAAGCGAATGCGGCTGCTCATGGAAGTGCAGAAAG 496

RESULT 6
US-09-252-991A-13627/c

Sequence 13627, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252, 991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13627
LENGTH: 3720
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13627

Query Match 5.0%; Score 69; DB 4; Length 3720;
Best Local Similarity 52.2%; Pred. No. 4,7e-08;
Matches 153; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 84 GCCCGATGTGGGCGCACTGGCAAGCTGGAAGGAAGATTGAATTGCGCTTCATCAA 143
DB 676 GGCGTGGCGCGCGGTTCCAGACCGGAAGAGGCGCTTGACATCGTTTCATGCG 617
QY 144 GCTACCGACATGGCGCGGCTGGCGCGCGCGGCGGAAAAGGCTTCTTCGAGACGAGGG 203
DB 616 GCTGACCGAATTCGCGCTGCTGATGCTGGCGCGGCGGCGGAGGCTTCGCCGCTTACGG 557
QY 204 CTTGTTCTGCAACTGGAAGCGGCAACTGGAAGGTGTGATGATGAGGTGCTGAA 263
DB 556 CTTACGCTCAATCTCAGGCGCGCAACCGTCTGCGGCCACCTTGCAGCAAGCTGCTCAG 497
QY 264 TGGCGAAGTGAAGCGCTGCGACATGCTGGCGCGCGCGCGCTTACCGCGCGAGCTTGGCTT 323
DB 496 CGGCGAGCTGAATGCCCGGAGTGCCTGTACGGGCTTGTACGGGCTTCAGGCTCGGCTT 437
QY 324 CGGCAACCAAGCGCGATATGAGGTGCGCTTACGATGAGGCTTCAACGCGCAAG 376
DB 436 CGGCGGAGCGCGCGCAAGCGAATGCGGCTGCTCATGGAAGTGCAGAAAG 384

RESULT 7
US-09-252-991A-9136

Sequence 9136, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252, 991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 9136
;; LENGTH: 1260
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9136

Query Match 4.2%; Score 58; DB 4; Length 1260;
Best Local Similarity 56.2%; Pred. No. 2,4e-05;
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

QY 131 TCAGCTTCATCAAGCTCACCAGATGAGCGCGCTGGCGGCGCGGAAAGGCTTCT 190
DB 197 TCAGCTTCATCAAGCTCACCAGATGAGCGCGCTGGCGGCGCGGAAAGGCTTCT 256
QY 191 TCAGAGACGAGGCGCTGTTCTGTCGCACTGAGAGCGGAGGCGC--AACTGAGAGTGTGA 247
DB 257 TCAGAGCGGAGGCGCTGTTCTGTCGCACTGAGAGCGGAGGCGC--AACTGAGAGTGTGA 316
QY 248 TGATAGGCTTCGTAATGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 307
DB 317 TCAGAGCGGCTTCATCTCCGCGCAGGTCAAGTCACTGCTGCTGCTGCTGCTGCTGCT 376
QY 308 CGGCGACGCTTTCGCGCAGGAGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 367
DB 377 GGGC-----CGCTACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 430
QY 368 ACGGCAACGCGATTCAGTGTCTCCATGAAATC 399
DB 431 GCGGCTCGGGGCTCAGCGTGTCTCCGAGATC 462

RESULT 8
US-09-328-352-3446
;; Sequence 3446, Application US/09328352
;; Patent No. 6562958
;; GENERAL INFORMATION:
;; APPLICANT: Gary L. Breton et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;; FILE REFERENCE: GTC99-03PA
;; CURRENT APPLICATION NUMBER: US/09/328,352
;; PRIORITY FILING DATE: 1999-06-04
;; NUMBER OF SEQ ID NOS: 8252
;; SEQ ID NO 3446
;; LENGTH: 1011
;; TYPE: DNA
;; ORGANISM: Acinetobacter baumannii
US-09-328-352-3446

Query Match 3.6%; Score 50; DB 4; Length 1011;
Best Local Similarity 51.8%; Pred. No. 0.0028;
Matches 113; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 658 CCGCGCGCGCAATGCGCTGCAAGCTGGAAGCGGAGCAATTTTCGTTATTCGCTGCGC 717
DB 496 CCGCGCGCGCAATGCGCTGCAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGA 555
QY 718 GAGCGCTGGAACGAGGCGCGCTTCAAGGCGATAGCGCGCTGATCAACCGATGAA 777
DB 556 GAGCGCTGGAATGATCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 615
QY 778 GAACTCTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 837
DB 616 GAACTATATCCCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 675
QY 838 TACCCCAACACTATATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 875
DB 676 CATCTCAGACCTTATGTTGCGCTTACAGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 713

RESULT 9

US-09-902-540-7805
;; Sequence 7805, Application US/09902540
;; Patent No. 6833447
;; GENERAL INFORMATION:
;; APPLICANT: Goldman, Barry S.
;; APPLICANT: Hinkle, Gregory J.
;; APPLICANT: Slater, Steven C.
;; APPLICANT: Wiegand, Roger C.
;; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;; FILE REFERENCE: 38-10(15849)B
;; CURRENT APPLICATION NUMBER: US/09/902,540
;; PRIORITY FILING DATE: 2001-07-10
;; PRIOR FILING DATE: 2000-07-10
;; NUMBER OF SEQ ID NOS: 16825
;; SEQ ID NO 7805
;; LENGTH: 1368
;; TYPE: DNA
;; ORGANISM: Myxococcus xanthus
US-09-902-540-7805

Query Match 3.5%; Score 48.2; DB 4; Length 1368;
Best Local Similarity 44.3%; Pred. No. 0.0095;
Matches 197; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

QY 195 GAGCGAGGCGCTTTCGTCGCACTGGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 254
DB 93 GGTCAAGTCAAGTTCATTAACCGGCGGCTTCATCAAGGAGGAGGAGGAGGAGGAGGAGG 152
QY 255 GGTGTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 314
DB 153 CCTCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 212
QY 315 CTTGCTTCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 374
DB 213 GTCGCGCAACACCGGAGTGGGCGTGGCGCTGGCGCGCTGCAAGGCTTACAGTGAT 272
QY 375 CCGGATTAAGTGTCCATGAATCTGAGTATGAAGCGGAGGAGGAGGAGGAGGAGGAGG 434
DB 273 CTTACGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 332
QY 435 CCGTAAACCGGTGATTCGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 494
DB 333 GAGAGTGTGTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 392
QY 495 AGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 554
DB 393 GCGCAAGCGCTGCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 452
QY 555 ACTGCTTATCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 614
DB 453 CGACCAATCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 512
QY 615 CATTTCCGCGCAATTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 639
DB 513 CAAGTTGACTACTTCTGTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 537

RESULT 10
US-09-902-540-777/C
;; Sequence 777, Application US/09902540
;; Patent No. 6833447
;; GENERAL INFORMATION:
;; APPLICANT: Goldman, Barry S.
;; APPLICANT: Hinkle, Gregory J.
;; APPLICANT: Slater, Steven C.
;; APPLICANT: Wiegand, Roger C.
;; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;; FILE REFERENCE: 38-10(15849)B
;; CURRENT APPLICATION NUMBER: US/09/902,540
;; PRIORITY FILING DATE: 2001-07-10
;; PRIOR FILING DATE: 2000-07-10

NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 777
LENGTH: 6655
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-777

Query Match 3.5%; Score 48.2; DB 4; Length 6655;
Best Local Similarity 44.3%; Pred. No. 0.017;
Matches 197; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

```
QY 195 GACGAGGCGCTTGTGTGCACTGGAAGCGCGCACTGGAAGGTGTGATGATAG 254
DB 4813 GGTCAAGTGAATGATGAAACCGGCGCGCTCATTAAGAACCGACGCGCTTACAT 4754
QY 255 GGTGTGTAATGCGCAACTGACCGCTTCGCAATGCTGCGCGCGCGCTTAAAGCGG 314
DB 4753 CCTCGAAGAGCCGAGCGGAGGGAAGCTCAAGCCCGCGGCGGACCATCGTCAAGAAC 4694
QY 315 CGTTGCGTTCGCAACCAAGCCGATATCGAGTGCCTTACAGATGCGCTTCAACGCGCA 374
DB 4693 GTCGCGCAACACCGCGCATGCGCGCTGCGCGCTGCGCGCGCTCAAGGCTTAAAGTGCAT 4634
QY 375 CGGATTAAGCGTGTCAATGAATATGCGCATGATGAAGCCGACATACCGCTGGAAG 434
DB 4633 CTTCACCATGCGCGCAAGATGTCCTTGAGAAAGATCAACCGCTCAAGCGCTGCGCGC 4574
QY 435 CGGTAAACCGGTGATCGATCAAGCAGATTAATCTTAAACCGGTGCTGCAAAATACAA 494
DB 4573 GCAAGTGTGTGATGACCGCGCAAGACGTGCGCGCGCGGACCGCGCAGCTACAGAGAC 4514
QY 495 AGCGGAAGGAGACCGGTCAATATGCGGATGACCTTCCGCGCGGATCTCAACATCAA 554
DB 4513 GCGCAAGCGCTGCAACCGCGAGAGACGCGCGCGCTTCACTCAACAGTACCAACCC 4454
QY 555 ACTGCGTATGCTGCGCGCTGCGGTATCAATCTGCTACTATTCGCGCGCGAGGA 614
DB 4453 CGACCAATGAGCGCGCATCAACACACCGGTCTGAGATCTCAAGAGACCGAGGG 4394
QY 615 CATTCGCGCAATCGCGCAGAC 639
DB 4393 CAAGTTCGACTACTTCTCGCGGCG 4369
```

RESULT 11

US-09-902-540-5834
Sequence 5834, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
PRIOR FILING DATE: 2001-07-10
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 5834
LENGTH: 750
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-5834

Query Match 3.4%; Score 48; DB 4; Length 750;
Best Local Similarity 47.9%; Pred. No. 0.0086;
Matches 138; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
QY 144 GCTCACCGAATGAGCGCGCTGCGGTGCGCGCGCAAAAGGCTTCTTCAAGACGAGGG 203
DB 204 GCGCGCGAAGTTCGCGAGGTGACGTGCTGTCAACACGCGGCGCTGCGGTG 263
```

```
QY 204 CTTGTTCTGCACTGGAAGCGCGCAACTGGAAGGTGTGATGATGAGGTGCGTGA 263
DB 264 GACCTGCGCAATCGACCCCGCTGAGAGCTGGAGACATGATGTGACCAACGAGAA 323
QY 264 TGGCAACTGACGCGCTCGACATGCTGCGCGCGCGCGCTTAAAGCGCGCTTGGCTT 323
DB 324 GGGCTCTCTTACTGACGACGCGCGGTGCTGCGCGCGCATGTGTGGCGGCAACCGGGCCA 383
QY 324 CGGCAACCAAGCGCGATATCGAGGTGCGCTTCAAGATGAGGCTTCAACGCGACCGATTAC 383
DB 384 GCTGTCAACATGAGGTGCTGCGTGGCGCGAGTGGCCCTACCGGCGGCAACGTGTACGG 443
QY 384 GGTGTCAATGAATCTGGATCAAGTGAAGCCGAACTACCGCTGGA 431
DB 444 CGCACCAAGCGCTTGTGTGACCAAGTTCAGCTTCAACCTCGCGCGGA 491
```

RESULT 12

US-09-902-540-1273/c
Sequence 1273, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
PRIOR FILING DATE: 2001-07-10
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1273
LENGTH: 72704
TYPE: DNA
ORGANISM: Myxococcus xanthus
FEATURE:
NAME/KEY: unsure
LOCATION: (1)-(72704)
OTHER INFORMATION: unsure at all n locations
US-09-902-540-1273

Query Match 3.4%; Score 48; DB 4; Length 72704;
Best Local Similarity 47.9%; Pred. No. 0.046;
Matches 138; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
QY 144 GCTCACCGAATGAGCGCGCTGCGGTGCGCGCGCAAAAGGCTTCTTCAAGACGAGGG 203
DB 70204 GCGCGCGAAGTTCGCGAGGTGACGTGCTGTCAACACGCGGCGCTGCGGTG 70145
QY 204 CTTGTTCTGCACTGGAAGCGCGCAACTGGAAGGTGTGATGATGAGGTGCGTGA 263
DB 70144 GACCTGCGCAATCGACCCCGCTGAGAGCTGGAGACATGATGTGACCAACGAGAA 70085
QY 264 TGGCAACTGACGCGCTCGACATGCTGCGCGCGCGCGCTTAAAGCGCGCTTGGCTT 323
DB 70084 GGGCTCTCTTACTGACGACGCGGTGCTGCGCGCGCATGTGTGGCGGCAACCGGGCCA 70025
QY 324 CGGCAACCAAGCGCGATATCGAGGTGCGCTTCAAGATGAGGCTTCAACGCGACCGATTAC 383
DB 70024 GCTGTCAACATGAGGTGCTGCGTGGCGGAGTGGCCCTACCGGCGGCAACGTGTACGG 69965
QY 384 GGTGTCAATGAATCTGGATCAAGTGAAGCCGAACTACCGCTGGA 431
DB 69964 CGCACCAAGCGCTTGTGTGACCAAGTTCAGCTTCAACCTCGCGCGGA 69917
```

RESULT 13

US-08-592-874-1/c
Sequence 1, Application US/08592874
Patent No. 5854034

```

GENERAL INFORMATION:
APPLICANT: POLLOCK, THOMAS J.
APPLICANT: YAMAZAKI, MOTOHIDE
APPLICANT: THORNE, LINDA
APPLICANT: MIKOLAJCZAK, MARCIA
APPLICANT: ARMENTROUT, RICHARD W.
TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS: 1
ADDRESSEE: JULES E. GOLDBERG
STREET: 261 MADISON AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,874
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/377,440
FILING DATE: 24-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: GOLDBERG, JULES E.
REGISTRATION NUMBER: 24,408
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-986-4090
TELEFAX: 212-818-9479
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 28804 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal
US-08-592-874-1

Query Match      3.4%; Score 47.8; DB 2; Length 28804;
Best Local Similarity 47.0%; Pred. No. 0.037;
Matches 148; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 64 GTTTGGGGTTTGAAGATTTGGGCGCCGATGTCGGGGCAGTTGGCGAAGTGAAGAAAGAAAGT 123
DB 18736 GTGGCGGGGCTGGCGATCAACACCGCTGGCTGGCAATCGCCGCTGGAAGTGTGGCG 18677
QY 124 TTGAATTCGGCTTCAAGCTCACCGACATGCGCCGCTGGCGTGGCCGCGGCGAATAA 183
DB 18676 CGAGCGCAGGGCGGATCGCCCGCGATCGGGAGACCAAGATGTGAGTCCCGGAAAGC 18617
QY 184 GCGTTTTCGAGAGCAGAGGCGCTGTTCTGTGCACTGGAAGCGCAGCCCACTGGAAGGTG 243
DB 18616 GGTATGTCGCGGCGCATTTCTGTGTGGCGCAGAGGCGCAAGAGGTTTCAAGAGGCGCAGGTGCTG 18557
QY 244 GTGATGATAGAGGTGCTGTAATGCGCACTGGAACGGCTGCCCAATGCTGCGCGCGCGCGCG 303
DB 18556 ATCAAGCTGTGATCCCACTGTCGGAGCCGGAAGCGCGCGAGCGCGCTGTGCTGTC 18497
QY 304 TTAGCGCCAGCGTTGGCTTCGCGACCAAGCGCGATATGAGGTGCGCTTTCAGCATGGGC 363
DB 18496 ACGCGCCAGCTGATGCCCGCGCGCAACAGCGCATATGACGCGCTGGAAGGAAAGGC 18437
QY 364 TTCAACGGCAACGCG 378
DB 18436 TTCGCTTCGCGCG 18422
```

```

RESULT 14
US-09-942-2/c
Sequence 2, Application US/09096942
Patent No. 6027925
GENERAL INFORMATION:
APPLICANT: Pollock, Thomas J
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
APPLICANT: Armentrout, Richard W
TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT APPLICATION NUMBER: US/09/096,942
CURRENT FILING DATE: 1998-06-12
EARLIER APPLICATION NUMBER: 60/049,428
EARLIER FILING DATE: 1997-06-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 28804
TYPE: DNA
ORGANISM: Sphingomonas sp. S88
US-09-942-2

Query Match      3.4%; Score 47.8; DB 3; Length 28804;
Best Local Similarity 47.0%; Pred. No. 0.037;
Matches 148; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 64 GTTTGGGGTTTGAAGATTTGGGCGCCGATGTCGGGGCAGTTGGCGAAGTGAAGAAAGAAAGT 123
DB 18736 GTGGCGGGGCTGGCGATCAACACCGCTGGCTGGCAATGCGCCGCTGGAAGTGTGGCG 18677
QY 124 TTGAATTCGGCTTCAAGCTCACCGACATGCGCCGCTGGCGTGGCCGCGGCGAATAA 183
DB 18676 CGAGCGCAGGGCGGATCGCCCGCGATCGGGAGACCAAGATGTGAGTCCCGGAAAGC 18617
QY 184 GCGTTTTCGAGAGCAGAGGCGCTGTTCTGTGCACTGGAAGCGCAGCCCACTGGAAGGTG 243
DB 18616 GGTATGTCGCGGCGCATTTCTGTGTGGCGCAGAGGCGCAAGAGTTCAGAAAGGCGCAGGTGCTG 18557
QY 244 GTGATGATAGAGGTGCTGTAATGCGCACTGGAACGGCTGCCCAATGCTGCGCGCGCGCGCG 303
DB 18556 ATCAAGCTGTGATCCCACTGTCGGAGCCGGAAGCGCGCGAGCGCGCTGTGCTGTC 18497
QY 304 TTAGCGCCAGCGTTGGCTTCGCGACCAAGCGCGATATGAGGTGCGCTTTCAGCATGGGC 363
DB 18496 ACGCGCCAGCTGATGCCCGCGCGCAACAGCGCATATGACGCGCTGGAAGGAAAGGC 18437
QY 364 TTCAACGGCAACGCG 378
DB 18436 TTCGCTTCGCGCG 18422

RESULT 15
US-09-942-2/c
Sequence 2, Application US/09096867
Patent No. 6030817
GENERAL INFORMATION:
APPLICANT: Pollock, Thomas J
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
APPLICANT: Armentrout, Richard W
TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT APPLICATION NUMBER: US/09/096,867
CURRENT FILING DATE: 1998-06-11
EARLIER APPLICATION NUMBER: 60/049,428
EARLIER FILING DATE: 1997-06-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
```

; SEQ ID NO 2
; LENGTH: 28804
; TYPE: DNA
; ORGANISM: *Sphingomonas* sp. S88
US-09-096-867-2

Query Match 3.4%; Score 47.8; DB 3; Length 28804;
Best Local Similarity 47.0%; Pred. No. 0.037;
Matches 148; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

```
QY 64 GTTGGGGTTTGAAGATTGGCCGATGTGGGGCAGTTGCAAGCTGAAAAAGAAAT 123
   |||||
DB 18736 GTGGCGGGGCTGGCGATCACACCGCTGGCTGGCAATCGGCCGTGAAATGTGTGCG 18677
   |||||
QY 124 TTGAAATTGGCTTTCATCAAGCTCACCGCATGCGCGCTGGCGGTGGCCGAAAAA 183
   |||||
DB 18676 CCGACGAGGGCGCGCATGCGCCCGATCGCGAGACCAAGATCGTGCATCCCCGAAAGC 18617
   |||||
QY 184 GCGTTCTTGAAGACGAGGGCTGTTGTGCAACTGGAAGCGCAAGCCAACTGGAAGTG 243
   |||||
DB 18616 GGTATGTCGCGCGCATTTCTGTGTGGGAGGGCGAAGGTTCAGAAGGGCCAGGTGCTG 18557
   |||||
QY 244 GTGATGATATGGGTCGTGAATTGGGAATGGACGCTCGACATGCTGGCGCGGCGCG 303
   |||||
DB 18556 ATCACGCTCGATCCACCATGTGGCGACCGAAGCCCGCAAGGCGCTGTGCGTGTCTC 18497
   |||||
QY 304 TTAGCGGCCAGCGTTGGCTTCGGCACCAAGCCGATATCGAGGTGCCGTTCAAGCATGGGC 363
   |||||
DB 18496 AGGCGCCAGCTCGATGCGCGCGCAACCAAGCGATCATGACGCGCTGAGCGAAAGGC 18437
   |||||
QY 364 TTCACGCGCAAGCG 378
   |||||
DB 18436 TTCGCTTCGCGCG 18422
   |||||
```

Search completed: May 20, 2005, 23:22:36
Job time : 261 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2005, 19:59:08 ; Search time 889 Seconds
(without alignments)
9602.015 Million cell updates/sec

Title: US-10-689-200-1

Perfect score: 1392
Sequence: 1 atgaacacatcatgatac.....ccgcgcgaagtcgtgat 1392

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 5695437 seqs, 3066160638 residues

Total number of hits satisfying chosen parameters: 11390874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: Published Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCRUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1392	100.0	1392	US-10-689-200-1	Sequence 1, Appl1
2	143	10.3	551	US-10-689-200-3	Sequence 3, Appl1
3	71.8	5.2	1209	US-10-282-122A-33791	Sequence 33791, A
4	62.2	4.5	1173	US-10-282-122A-11998	Sequence 11998, A
5	59.2	4.3	576	US-10-282-122A-11492	Sequence 11492, A
6	58	4.2	1200	US-10-282-122A-30186	Sequence 30186, A
7	58	4.2	1200	US-10-389-647-149	Sequence 149, App
8	57	4.1	1116	US-10-282-122A-33179	Sequence 32179, A
9	49.4	3.5	1512	US-10-369-493-42267	Sequence 42267, A
10	48.4	3.5	773	US-10-282-122A-8343	Sequence 8343, Ap
11	48	3.4	744	US-10-369-493-43243	Sequence 43243, A

12	47.8	3.4	1017	US-10-437-963-85118	Sequence 85118, A
13	47.8	3.4	7656	US-10-282-122A-25488	Sequence 25488, A
14	46.4	3.3	3048	US-10-156-761-3146	Sequence 3146, Ap
15	46.4	3.3	9025608	US-10-156-761-1	Sequence 1, Appl1
16	46.2	3.3	1440	US-10-437-963-82648	Sequence 82648, A
17	45.6	3.3	1173	US-10-282-122A-13773	Sequence 13773, A
18	45.4	3.3	1396	US-10-767-701-13423	Sequence 13423, A
19	44.4	3.2	930	US-10-437-963-93713	Sequence 93713, A
20	44.4	3.2	1365	US-10-437-963-55913	Sequence 55913, A
21	44	3.2	786	US-10-437-963-13840	Sequence 13840, A
22	43.4	3.1	1188	US-10-282-122A-13623	Sequence 13623, A
23	43.4	3.1	5577	US-10-369-493-43212	Sequence 43212, A
24	43.2	3.1	1155	US-10-437-963-67845	Sequence 67845, A
25	43.2	3.1	2110	US-10-437-963-10628	Sequence 10628, A
26	43	3.1	1305	US-10-282-122A-11478	Sequence 11478, A
27	43	3.1	1813	US-10-369-493-27231	Sequence 27231, A
28	43	3.1	2021	US-10-437-963-49750	Sequence 49750, A
29	42.8	3.1	956	US-10-425-114-24149	Sequence 24149, A
30	42.8	3.1	984	US-10-425-114-118414	Sequence 118414, A
31	42.8	3.1	993	US-10-437-963-93714	Sequence 93714, A
32	42.6	3.1	2213	US-10-772-636-27	Sequence 27, Appl1
33	42.6	3.1	2931	US-10-282-122A-15175	Sequence 15175, A
34	42.4	3.0	1132	US-10-156-761-4985	Sequence 4985, Ap
35	42.4	3.0	1305	US-10-282-122A-12878	Sequence 12878, A
36	42.4	3.0	1470	US-09-815-242-7855	Sequence 7855, Ap
37	42.4	3.0	9025608	US-10-156-761-1	Sequence 1, Appl1
38	42.2	3.0	375	US-10-425-114-4691	Sequence 4691, App
39	42.2	3.0	501	US-10-425-114-319	Sequence 319, App
40	42.2	3.0	726	US-10-425-114-58	Sequence 58, Appl1
41	42.2	3.0	1525	US-10-369-493-27882	Sequence 27882, A
42	42.2	3.0	1866	US-10-425-114-74324	Sequence 74324, A
43	42	3.0	536	US-10-338-110-119	Sequence 119, App
44	41.6	3.0	900	US-10-369-493-41628	Sequence 41628, A
45	41.6	3.0	1271	US-10-437-963-38282	Sequence 38282, A

ALIGNMENTS

RESULT 1
US-10-689-200-1
; Sequence 1, Application US/10689200
; Publication No. US20040126848A1
; GENERAL INFORMATION:
; APPLICANT: Dicostimo, Deana J.
; APPLICANT: Ni, Hao
; APPLICANT: Ye, Rick
; APPLICANT: Picaesgio, Stephen
; APPLICANT: Wang, Tao
; APPLICANT: Seid, John E.
; TITLE OR INVENTION: NATURAL PROMOTERS FOR GENE EXPRESSION IN CL METABOLIZING BACTERIA
; FILE REFERENCE: C11747 US NA
; CURRENT APPLICATION NUMBER: US/10/689,200
; PRIOR FILING DATE: 2003-10-20
; PRIOR FILING DATE: 2002-10-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1392
; TYPE: DNA
; ORGANISM: *Mechyliomonas* sp. 16a
US-10-689-200-1

Query Match 100.0%; Score 1392; DB 18; Length 1392;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAACCATCTTGTGATCGAGCTGAAGAAATTTGTTATGACGTTATCGGCTGCTA 60
DB 1 ATGAAACCATCTTGTGATCGAGCTGAAGAAATTTGTTATGACGTTATCGGCTGCTA 60
CY 61 GCCGTTGGGTTTGACGATTCGCGCCGATGTCGGGGCAGTTGGCAAGCTGAAAGGAA 120

Db 61 GCGGTTGGGGTTGAGGATTTGGCCGATGTGCGGGCACTTGGCAAGCTGGAAAAAGAA 120
Qy 121 GATTGAAATTCGAGCTTCATCAAGCTCAAGCAATGGCGCGCTGGCGATGGCGCGGAA 180
Db 121 GATTGAAATTCGAGCTTCATCAAGCTCAAGCAATGGCGCGCTGGCGATGGCGCGGAA 180
Qy 181 AAGGCTCTTCTGAGGAGCGAGGGCGTGTGTCGCAACTGGAAACGCGAGCGCACTGGAA 240
Db 181 AAGGCTCTTCTGAGGAGCGAGGGCGTGTGTCGCAACTGGAAACGCGAGCGCACTGGAA 240
Qy 241 GTGGTATGATAGGATGGCTGTAATGGCGAATGGCAAGCTCGCACTGTGGCGCGCGG 300
Db 241 GTGGTATGATAGGATGGCTGTAATGGCGAATGGCAAGCTCGCACTGTGGCGCGCGG 300
Qy 301 CCGTTAGCGGCGCAGCGTTGGCTTGGCGCAAGAGCGGATTCGAGGTGCGTTCAAGATG 360
Db 301 CCGTTAGCGGCGCAGCGTTGGCTTGGCGCAAGAGCGGATTCGAGGTGCGTTCAAGATG 360
Qy 361 GCGTTCAAGCGGCAAGCGGATTCGAGGTGCGCAAGATTCGAGGTGCGGAGCGGAGC 420
Db 361 GCGTTCAAGCGGCAAGCGGATTCGAGGTGCGCAAGATTCGAGGTGCGGAGCGGAGC 420
Qy 421 ATACCGCTGGAAGCGCGTAAACCGGTGCAATCCGATCAAGGAGGATTAATCAAAACCG 480
Db 421 ATACCGCTGGAAGCGCGTAAACCGGTGCAATCCGATCAAGGAGGATTAATCAAAACCG 480
Qy 481 GTGCAAAAATACAAAGCGGAGGCAAGCGGTCATATGCGATGACCTTCCGCGCGGA 540
Db 481 GTGCAAAAATACAAAGCGGAGGCAAGCGGTCATATGCGATGACCTTCCGCGCGGA 540
Qy 541 TCTCAAAATCAAAATGCTGTTCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 541 TCTCAAAATCAAAATGCTGTTCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCT 600
Qy 601 TCGCGCGCGGAGGCAATTCGCGCAAAATCGGCGGAGAGCGCTTGTGTGCGTGAACCG 660
Db 601 TCGCGCGCGGAGGCAATTCGCGCAAAATCGGCGGAGAGCGCTTGTGTGCGTGAACCG 660
Qy 661 CCGCGCGGAAATGCGTCAAGCTGCAAGCGGAGGCAATTCGCTTATGCGTGGCGAG 720
Db 661 CCGCGCGGAAATGCGTCAAGCTGCAAGCGGAGGCAATTCGCTTATGCGTGGCGAG 720
Qy 721 CCGTGAACGAGAGGCGGCTTCAAGGCGATGAGGCGGCTGCTGCTGCTGCTGCTGCT 780
Db 721 CCGTGAACGAGAGGCGGCTTCAAGGCGATGAGGCGGCTGCTGCTGCTGCTGCTGCT 780
Qy 781 CTCTGGAAGGACGCGCGGAAAAAGTCTTGGCGTGAACCAATGAGGCGGAGAAATAC 840
Db 781 CTCTGGAAGGACGCGCGGAAAAAGTCTTGGCGTGAACCAATGAGGCGGAGAAATAC 840
Qy 841 CCCAACAATCTATCTGCGGCTGCAAGGCGCTGATTCGCGGCGCGATTCGCTGAGCG 900
Db 841 CCCAACAATCTATCTGCGGCTGCAAGGCGCTGATTCGCGGCGCGATTCGCTGAGCG 900
Qy 901 GACAAATTAAGAAAGCGCAAGGAGGCAATGCAATGCTGCGGCAAAATCAATTCGCT 960
Db 901 GACAAATTAAGAAAGCGCAAGGAGGCAATGCAATGCTGCGGCAAAATCAATTCGCT 960
Qy 961 GGTGAGTGAAGTGTGCGGCTGATGCAATGAGGCGGCTTGGATTCGCAATTCGCAAT 1020
Db 961 GGTGAGTGAAGTGTGCGGCTGATGCAATGAGGCGGCTTGGATTCGCAATTCGCAAT 1020
Qy 1021 AAAAGCGGCTACCGGACTTCAACACCTTCTTTCGCGACGCGCGAGCTATCGCTCTAC 1080
Db 1021 AAAAGCGGCTACCGGACTTCAACACCTTCTTTCGCGACGCGCGAGCTATCGCTCTAC 1080
Qy 1081 AGCAGTGAAGTGTGCTGATTCGACCCGAGGCGGCTGAGGCGGCTGAGTTCATGATTC 1140
Db 1081 AGCAGTGAAGTGTGCTGATTCGACCCGAGGCGGCTGAGGCGGCTGAGTTCATGATTC 1140
Qy 1141 CCGGACAACTGGATTCGATCCGCGCAAGAGCTTACCGCGCGGACATCTATCGCGC 1200
Db 1141 CCGGACAACTGGATTCGATCCGCGCAAGAGCTTACCGCGCGGACATCTATCGCGC 1200

Db 1141 CCGGACAACTGGATTCGATCCGCGCAAGAGCTTACCGCGCGGACATCTATCGCGC 1200
Qy 1201 GCGGCCAAGAACTGGTTCGAGAGGCAAGGCCAAGGCGGAGACTTCCCTGCGATACC 1260
Db 1201 GCGGCCAAGAACTGGTTCGAGAGGCAAGGCCAAGGCGGAGACTTCCCTGCGATACC 1260
Qy 1261 TCGATCAAGCCGTCGAGAAATTTCTTCAATGCAAAAGTCCGTTGATGCAACAGGCC 1320
Db 1261 TCGATCAAGCCGTCGAGAAATTTCTTCAATGCAAAAGTCCGTTGATGCAACAGGCC 1320
Qy 1321 AACGATTACCTGCGCAAGTTTGGATAGGTGTAAGGCAAGCAACCGTAGCGCGGCG 1380
Db 1321 AACGATTACCTGCGCAAGTTTGGATAGGTGTAAGGCAAGCAACCGTAGCGCGGCG 1380
Qy 1381 AAGTCTGTGAT 1392
Db 1381 AAGTCTGTGAT 1392

RESULT 2
US-10-689-200-3
; Sequence 3, Application US/10689200
; Publication No. US20040126848A1
; GENERAL INFORMATION:
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Ni, Hao
; APPLICANT: Ye, Rick
; APPLICANT: Picatagallo, Stephen
; APPLICANT: Wang, Tao
; APPLICANT: Seip, John E.
; TITLE OF INVENTION: NATURAL PROMOTERS FOR GENE EXPRESSION IN CL METABOLIZING BACTERIA
; FILE REFERENCE: CL1747 US NA
; CURRENT APPLICATION NUMBER: US/10/689,200
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: 60/419,872
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Methylobionas sp. 16a
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (409)..(411)
; OTHER INFORMATION: ATG start site of nra gene
US-10-689-200-3

Query Match 10.3%; Score 143; DB 18; Length 551;
Best Local Similarity 100.0%; Pred. No. 7.1e-34;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAAACCATATAGATGAGCTCGAAGAAATTTGTTATGACGTTATCGGCTTCGCTA 60
Db 409 ATGAAAAACCATATAGATGAGCTCGAAGAAATTTGTTATGACGTTATCGGCTTCGCTA 468
Qy 61 GCGGTTGGGGTTTGAAGATTGCGCCCGATGTCGCGGCGAGTTGGCGAAGTGAAGAAAGAA 120
Db 469 GCGGTTGGGGTTTGAAGATTGCGCCCGATGTCGCGGCGAGTTGGCGAAGTGAAGAAAGAA 528
Qy 121 GATTGAAATTCGCTTATCA 143
Db 529 GATTGAAATTCGCTTATCA 551

RESULT 3
US-10-282-122A-33791
; Sequence 33791, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangou
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIORITY FILING DATE: 2003-02-20
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/230,335
PRIORITY FILING DATE: 2000-09-06
PRIORITY APPLICATION NUMBER: 60/230,347
PRIORITY FILING DATE: 2000-09-09
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/267,636
PRIORITY FILING DATE: 2001-02-09
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33791
LENGTH: 1209
TYPE: DNA
ORGANISM: Pseudomonas syringae
US-10-282-122A-33791

Query Match 5.2%; Score 71.8; DB 17; Length 1209;
Best Local Similarity 53.8%; Pred. No. 2e-11;
Matches 148; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 102 TGGCAAGCTGGAAGAAAGATTGAAATTCGGCTTCACTCAAGCTCACCGACATGCGCC 161
DB 45 TACCGCCCGGAAATGAACAGCCCTGACGTGGGCTTCAATGCACTGACGACCTGCGCCC 104
QY 162 GCTGCGCGTGGCGCGCGAAAGGCTTCTCGAGGACGAGGGCGCTGTTCTGCAACTGGA 221
DB 105 GCTGCTGTGTGCTGCTCCACCCAGGGGTTTGCCCACTTACGGGCTTGAGCTTGAACCTGAA 164
QY 222 AGCGCAGGCACTGGAAGGTGTGATGATAGGCTGTGATAGGCGCAACTGGAAGCTGCA 281
DB 165 ACCGCAAGACTCTCTGGGCGCGGCTGCGTGAAGGCTGTCAAGCGGCGAGCTTCAAGCGGCG 224
QY 282 GCACATGCTGTGGCGCGCGCGCTTGAAGCGGCGAGGCTTGGCTTGGCGCAAGCGCGATAT 341
DB 225 AACTAGGCTGTATAGTGTGATCTATAGCTGTGAGCTGGGCAATCAAGCGCGCGCGGCGAC 284
QY 342 CGAGTGGCGCTTCAAGATGAGGCTTCAAGCGGAAGC 376
DB 285 CGACATGGCACTCTCATGAGGCTTCAACGAAAGC 319

RESULT 4
US-10-282-122A-11998
Sequence 11998, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIORITY FILING DATE: 2003-02-20
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/230,335
PRIORITY FILING DATE: 2000-09-06
PRIORITY APPLICATION NUMBER: 60/230,347
PRIORITY FILING DATE: 2000-09-09
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/267,636
PRIORITY FILING DATE: 2001-02-09
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11998
LENGTH: 1173
TYPE: DNA
ORGANISM: Burkholderia cepacia
US-10-282-122A-11998

Query Match 4.5%; Score 62.2; DB 17; Length 1173;
Best Local Similarity 58.4%; Pred. No. 2e-08;
Matches 128; Conservative 0; Mismatches 88; Indels 3; Gaps 1;
QY 131 TCGGCTTCACTCAAGCTCAGACATGCGCGCTGCGGTGCGCGCGGAAAGGCTTCT 190
DB 116 TCGGCTATCTGCGGATCAGACGCGGCGCGCTGCTGTGTCGCCACACAAACGCTATAT 175
QY 191 TCGAGACGAGGCGCTGT---TGTGCACTGGAAGCGCAGGCCCACTGGAAGTGTGA 247
DB 176 TCGGCTCCGAAAGGCTGACGCTGACGACCAAGAGCTGTGCTGAGTGGGCGGAGCTCG 235
QY 248 TGGATAGGCTCGTGAATGCGGAATGGAAGCTCGACATGCTGAGCGCGCGCGCTTAG 307
DB 236 TCGAGGCGTTCTGTGTCCGACAGGTCAACGCTGTGCACTGCTGTGCGCATGACGCTGT 295
QY 308 CGGCAAGCGCTTGGCTTCCGACCAAGCGCGCATATCGAGG 346
DB 296 GGGGCGGCTACGCGCAGCGCGCGCGCGGCAAGGTGTGTG 334

RESULT 5
US-10-282-122A-11492
Sequence 11492, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11492
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-282-122A-11492

```

```

Query Match      4.3%; Score 59.2; DB 17; Length 576;
Best Local Similarity 57.9%; Pred. No. 1.3e-07;
Matches 155; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

QY 131 TCGGCTTCATCAAGCTCACCAGCATGCGCGCTGCGGTGCGCGCGGAAAAGGCTTCT 190
    |||||
DB 116 TCGGCTATCTGCGGATCAACGACGCGCGCTGCTGTGCTGCCCAACACGCTATT 175
    |||||
QY 191 TCGAGACGAGGCGCTGT--TGTGCAACTGGAAGCGCAAGCCAACTGGAAAGTGTGA 247
    |||||
DB 176 TCGCGCTCCAGAGGCTGACGCTGACCAACCCAACTGCTGCGCACTGGGCGCAACTG 235
    |||||
QY 248 TGGATAGGGTCTGATGATGAGCACTGAGCGGCTGCACTGCGGCGCGCGCTTAG 307
    |||||
DB 236 TCGAGCGCTTCTCTGCTCGGCGCAAGTCAAGCTGTGCACTGCTGCTGCGGATGACGCTGT 295
    |||||
QY 308 CGGCGAGCGTTGGCTTGGCAGCCAGCCAGCCGATATCG 343
    |||||
DB 296 GGGCAGCGTACGCGCAAGCCGCGCGCGCGCAAGGTG 331
    |||||

```

```

RESULT 6
US-10-282-122A-30186
; Sequence 30186, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari

```

```

; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 30186
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30186

```

```

Query Match      4.2%; Score 58; DB 17; Length 1200;
Best Local Similarity 56.2%; Pred. No. 4.1e-07;
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

QY 131 TCGGCTTCATCAAGCTCACCAGCATGCGCGCTGCGGTGCGCGCGGAAAAGGCTTCT 190
    |||||
DB 137 TCGGCTACTGCGGATCAACGACGCGCGCTTGTGCTGCGCCATGCAACGCGCTGT 196
    |||||
QY 191 TCGAGACGAGGCGCTTGTGCTGCAACTGGAAGCGCAGGCC--AACTGGAAGTGTGA 247
    |||||
DB 197 TCGAGGCTCCAGGCGCATCCAGCGCGCAACGCGCGCTGCTGTGCTGAGCTGGGCGCAAGTGA 256
    |||||
QY 248 TGGATAGGGTCTGATGATGAGCAACTGAGCGGCTGCACTGCGAGCGCGCGCTTAG 307
    |||||
DB 257 TCGAGCGCTTCTCTGCTCGGCGCAAGTCAAGCTATCACTGCTGTGCGGATGACCGTCT 316
    |||||
QY 308 CGGCGAGCGTTGGCTTGGCAGCCAGCCAGCCGATATGAGGTGCGCTTCAAGATGGCTTGA 367
    |||||
DB 317 GGGCC-----CGTACGCGCAAGAGTGC CGGCCAAGAGTGTGCTGCAACCAAGTGC 370
    |||||
QY 368 AGGCAAGCGGTTAAGGTGTTCATTAATTC 399
    |||||
DB 371 GCGGCTCGGGGCTCAAGGTGCTCGCGAGATC 402
    |||||

```

```

RESULT 7
US-10-389-647-149
; Sequence 149, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, B. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Candi

```

;; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
;; FILE REFERENCE: UIZ-038CP
;; CURRENT APPLICATION NUMBER: US/10/389,647
;; CURRENT FILING DATE: 2003-03-14
;; PRIOR APPLICATION NUMBER: 09/653730
;; PRIOR FILING DATE: 2000-09-01
;; PRIOR APPLICATION NUMBER: 60/153022
;; PRIOR FILING DATE: 1999-09-03
;; NUMBER OF SEQ ID NOS: 710
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 149
;; LENGTH: 1200
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-149

Query Match 4.2%; Score 58; DB 17; Length 1200;
Best Local Similarity 56.2%; Pred. No. 4.1e-07;
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

Qy 131 TCGGCTTCATCAAGCTCACCGCATGCGCGCTGGCGCGCGCGCAAAAAGCTTCT 190
Db 137 TCGGCTACCTCGCGATCACCGACGCCACCGCTTCTGCTGGCCCATGCGGCTGT 196
Qy 191 TCGAGACGAGGCGCTGTTCTGTCGCACTGAGACGCGAGGCC--AACTGAGAGTGTGTA 247
Db 197 TCGAGGCCGAGGCGATTCAGGCGGAAACGCGGCTGCTTGGTTCGAGTGGCGCAAGTGA 256
Qy 248 TGGATAGGCTGCGAATGCGGAACTGGAACGCTTCGACATGCTGGCGCGCGCGCTTAG 307
Db 257 TCGAGGCGTTTCATCTCCGGCAGGTCAACGTCATCCACCTGCTGCGCCCATGACCTCT 316
Qy 308 CGGCGACGCTTGGCTTCGCGACCAAGCGCGATATCGAGGTGCGCTTCAGCATGGGCTTCA 367
Db 317 GGGCC-----CGCTACGCGACGAGGTGCGCGCGCAAGGTGTGGCTGGAACACGCTCG 370
Qy 368 ACGGCAACGCGATTACGCTGTCTCCATGAAATC 399
Db 371 GCGGCTCGGGGCTCACGCTGTGCTCCGAGATC 402

RESULT 8
US-10-282-122A-32179
;; Sequence 32179, Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Liangu
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsyth, R.
;; APPLICANT: Xu, H.
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 32179
;; LENGTH: 1116
;; TYPE: DNA
;; ORGANISM: Pseudomonas putida
US-10-282-122A-32179

Query Match 4.1%; Score 57; DB 17; Length 1116;
Best Local Similarity 54.1%; Pred. No. 8.2e-07;
Matches 139; Conservative 0; Mismatches 115; Indels 3; Gaps 1;

Qy 131 TCGGCTTCATCAAGCTCACCGCATGCGCGCTGGCGCGCGCGCAAAAAGCTTCT 190
Db 56 TCGGCTACCTCGCGATCACCGATGCCACCGCTTCTGCTGGCGCATMAACGCGCTGT 115
Qy 191 TCGAGACGAGGCGCT---GTTGTCGCACTGGAAGCGCAGGCCAACTGGAAGTGTGTA 247
Db 116 TCGAGGCCGAGGCGATTCAGGCGGAAACGCGGCTGCTGCTGGCGAGTGGCGCAAGTGA 175
Qy 248 TGGATAGGCTGCGAATGCGGAACTGGAACGCTTCGCAATGCTGGCGCGCGCGCTTAG 307
Db 176 TCGAGCATTCATCTCCGGCAGGTCAACGTCATCCACCTGCTTGGCCCATGACGCTAT 235
Qy 308 CGGCGACGCTTGGCTTCGCGACCAAGCGCGATATCGAGGTGCGCTTCAGCATGGGCTTCA 367
Db 317 GGGCC-----CGCTACGCGACGAGGTGCGCGCGCAAGGTGTGGCTGGAACACGCTGGCGGCTT 295
Qy 368 ACGGCAACGCGATTACG 384
Db 296 GCGGCTCGACTGTTGCG 312

RESULT 9
US-10-369-493-42267
;; Sequence 42267, Application US/10369493
;; Publication No. US20030233675A1
;; GENERAL INFORMATION:
;; APPLICANT: Cao, Yongwei
;; APPLICANT: Hinkle, Gregory J.
;; APPLICANT: Slater, Steven C.
;; APPLICANT: Goldman, Barry S.
;; APPLICANT: Chen, Xianning
;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; FILE REFERENCE: 38-10(52052)B
;; CURRENT APPLICATION NUMBER: US/10/369,493
;; CURRENT FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 42267
;; LENGTH: 1512
;; TYPE: DNA
;; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-42267

Query Match 3.5%; Score 49.4; DB 17; Length 1512;
Best Local Similarity 47.8%; Pred. No. 0.00022;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
Qy 132 CGGCTTCATCAAGCTCACCGCATGCGCGCTGGCGCGCGCGCAAAAAGGCTTCT 191

Db 669 CGGCTCCGCGAGGTGGGAGAGGTCCAGCGGAGGTGCGCCGACGACCCCGGCAAGCT 728
Qy 192 CGAGAGCAGAGGCTGTGTGCACTGCAAGCGCGCACTGGAAGGTGTGATGA 251
Db 729 CGAGGCTCGGAGATGGGCGGCAAGACGCGATCGTGTCAACGAGGCGGACCTCA 788
Qy 252 TGGGCTGGAATGGGAACTGGACGCTTGGACATGTGGCGCGCGGCTGTACCGG 311
Db 789 CATGCCGTTACGCGCGGTCAATGAGCTGTTCAGACGACCGCGCGGCTGTGTTC 848
Qy 312 CAGCGTTGCTTGGACGCAAGCGCGATATCGAGTGGCGGTTGAGCATGGCTTCAACG 371
Db 849 CAGCGAGCGGCTATGTGTCAAGAGGACCTTCAAGAGATTCAACAGCGGTTGCTCA 908
Qy 372 CAGCGGATTACGCTGTCCATGAAATCTGGCATCTGATGAGCGGCAATACCGCTG 430
Db 909 GATCGCGGAGTGGTGTGACATGACGACCCCTTCAAGAGAGACGTTATGGGCGG 967

RESULT 10

US-10-282-122A-8343
; Sequence 8343, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PAM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 8343
; LENGTH: 773
; TYPE: DNA
; ORGANISM: *Acinetobacter baumannii*
US-10-282-122A-8343

Query Match 3.5%; Score 48.4; DB 17; Length 773;
Best Local Similarity 51.4%; Pred. No. 0.00035;
Matches 112; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 658 CCGCCGCGCAATATGCCCTCACGCTGGAAGCGGACACATTTGGTTATTCGCGG 717
Db 325 CCGCACCCCTATATGTGTCAAGCTTTGATTAACATGATGATGATGTTTGTGGA 384
Qy 718 GAGCCGTGGAACGAGGCGCGGTTCAGAGGCAATAGCGCGGTGATCAACGATGA 777
Db 385 GAGCCTTGGAATATCTAAGTGAATTTAGTTTAAGTAATAATGCTCTGCTCAATCA 444
Qy 778 GAATCTGGAAGGACAGCGGAAAAAGTCTTGCGGTGACCAACATAGGCGGGA 837
Db 445 GACATATTCCTCAATGTGCAATTAAGTTCTGCAATCAAGATAGGCGTGAACA 504
Qy 838 TACCCCAACCTATGTGGGGTGAACCAAGGGCTGAT 875
Db 505 CATCTCAACCTTATGTCCTTACACGGAATAT 542

RESULT 11

US-10-369-493-43243
; Sequence 43243, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(5205)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO: 43243
; LENGTH: 744
; TYPE: DNA
; ORGANISM: *Myxococcus xanthus*
US-10-369-493-43243

Query Match 3.4%; Score 48; DB 17; Length 744;
Best Local Similarity 47.9%; Pred. No. 0.00046;
Matches 138; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

Qy 144 GCTCACGACATGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 203
Db 204 GCGGCGGAGTTCGCGGAGGTGAGAGTGTGTCACAAAGGCGGCGGCGGCGGCTT 263
Qy 204 CCTGTTCTGCACTGGAAGCGGCACTGGAAGGTGTGATGATGATGATGATGAT 263
Db 264 GAGCTGCGGCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 323
Qy 264 TGGCGAATGAGCGGCTGCAATGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCTT 323
Db 324 GGGCTCTCTGTATGACCAAGCGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCA 383
Qy 324 CCGCACCAAGCGGATATGAGAGTGTGTCAGATGAGGCTTCAACGCGGCAATTTAC 383
Db 384 CGTGTCAACATGAGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 443
Qy 384 GGTGTCAATGAATCTGCGCATCAATGAGCGGCAACATACCGCTGGA 431
Db 444 CCGCACCAAGGCGCTTGTGCAACAGTTCAGCTCAACCTCCGCGGGA 491

RESULT 12

US-10-437-963-85118
; Sequence 85118, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.

Query Match	3.4%;	Score 47.8;	DB 18;	Length 1017;
Best Local Similarity	49.8%;	Pred. No. 0.0006;		
Matches 121;	Conservative 0;	Mismatches 122;	Indels 0;	Gaps 0.

RESULT 13
US-10-282-122A-25488

;; Sequence 255488, Application US/102821222A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:

APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zysekind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EUTPA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335

Query Match	3.4%;	Score 47.8;	DB 17;	Length 7656;
Best Local Similarity	54.2%;	Pred. No. 0.0014;		
Matches 97;	Conservative	0;	Mismatches 82;	Indels 0;
				Gaps 0;

RESULT 14
US-10-156-761-3146

; sequence 3146, Application US/10156/6
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ADDRESS: 10156/6

;
; APPLICANT: OMURA, SATOSHI
;
; APPLICANT: IKEDA, HARUO
;
; APPLICANT: ISHIKAWA, JUN
;
; APPLICANT: ISHIMURA, MASAHITO

; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262

;
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ;
 ; CURRENT FILING DATE: 2002-05-29
 ;
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ;

;
; PRIOR FILING DATE: 2001-05-30
;
; PRIOR APPLICATION NUMBER: JP 2001-272697
;
; PRIOR FILING DATE: 2001-08-02
;

```

; NUMBER OF SEQ ID NOS: 15109
;
; SEQ ID NO 3146
;
; LENGTH: 3048
;

```

```

; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: (1)..(3048)
US-10-156-761-3146

```

Query Match	3.3%	Score	46.4	DB	15	Length	3048
Best Local Similarity	50.0%	Pred. No.	0.0025				
Matches	116	Conservative	0	Mismatches	116	Indels	0
						Gaps	0

```
OY 69 GGGTTGACGATTCGCCCGATGTCGGGCGCATGTTGGCAAGCTGGAAAAGAGATTGAA 128
DB 645 GGTGCTGACGACCGCCCTGACGTCGTGGCCCTACACCGCGGAGAGACATCCC 704
OY 129 ATTGGCTTCATCAAGCTCACCGACATGGCGCGCTGGCGGTGGCCGGAAGAGCTT 188
DB 705 GCTCGGCATCCCATGTGGCCGCAAGAGCCCGAAGCTGAGAGGTGTGGCTGTT 764
OY 189 CTTGAGAGAGAGGCGCTGTTCGTGCACTGGAAGCGCAGGCCAATGGAAGGTGTGAT 248
DB 765 CATCAACATGCTGTGCTGGCACCGCATGTCTCGGCGACCCCACTTCGCCGAGCTGAT #24
OY 249 GGATAGGATCGTGAATGGAGCACTGGAGCGCTGGCAATGTCGGCGCGCG 300
DB 825 CGAGCGGATCGCGAGCGCCCAATGAGAGCTGTGACAGACACAGAGAGCTGCCG 876
```

RESULT 15

```
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

Query Match 3.3%; Score 46.4; DB 15; Length 9025608;

Best Local Similarity 50.0%; Pred. No. 0.07; Mismatches 116; Indels 0; Gaps 0;

```
OY 69 GGGTTGACGATTCGCCCGATGTCGGGCGCATGTTGGCAAGCTGGAAAAGAGATTGAA 128
DB 3936846 GGTGCTGACGACCGCCCTGACGTCGTGGCCCTACACCGCGGAGAGACATCCC 3936787
OY 129 ATTGGCTTCATCAAGCTCACCGACATGGCGCGCTGGCGGTGGCCGGAAGAGCTT 188
DB 3936786 GCTCGGCATCCCATGTGGCCGCAAGAGCCCGAAGCTGAGAGGTGTGGCTGTT 3936727
OY 189 CTTGAGAGAGAGGCGCTGTTCGTGCACTGGAAGCGCAGGCCAATGGAAGGTGTGAT 248
DB 3936726 CATCAACATGCTGTGCTGGCACCGCATGTCTCGGCGACCCCACTTCGCCGAGCTGAT 3936667
OY 249 GGATAGGATCGTGAATGGAGCACTGGAGCGCTGGCAATGTCGGCGCGCG 300
DB 3936666 CGAGCGGATCGCGAGCGCCCAATGAGAGCTGTGACAGACACAGAGAGCTGCCG 3936615
```

Search completed: May 20, 2005, 23:37:39
Job time : 899 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2005, 13:35:03 ; Search time 4942 Seconds
(without alignments)
10721.462 Million cell updates/sec

Title: US-10-689-200-1

Perfect score: 1392
Sequence: 1 atgaaacacatcatgacgc.....ccggcggaagtcgtgac 1392

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gsa1.*
9: gb_gsa2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93.4	6.7	1872	9	CL982367
2	87.6	6.3	575	8	A2935061 BJ_Ba000
3	82.4	5.9	800	8	B2575941 msh2_470.
4	61.8	4.4	406	7	CP161028 B0692703-
5	58.6	4.2	1085	8	B2548675 pacsl-60
6	49.2	3.5	1035	8	B2563077 pacsl-164
7	48.6	3.5	1407	8	B2572739 msh2_2775
8	48.4	3.5	788	8	B2550725 pacsl-60
9	47.8	3.4	844	9	AL056552 Drosophila
10	47.2	3.4	459	6	CA718866 wkm2n.pk0
11	47.2	3.4	691	6	CD894397 G118.126C
12	47.2	3.4	824	7	CN129290 RH01_34
13	46	3.3	623	6	CA699640 w1k8.pk0
14	45.6	3.3	469	6	CA066446 SCROAD101
15	45.6	3.3	500	6	CA069366 SCROAD106
16	45.6	3.3	602	6	CA069388 SCROAD106
17	45.6	3.3	616	6	CA219519 SCROAD106
18	45.6	3.3	645	6	CA212608 SCROAD111
19	45.6	3.3	693	6	CA264998 SCARTR304
20	45.6	3.3	708	6	CA069230 SCROAD105
21	45.6	3.3	758	6	CA212728 SCROAD112
22	45.6	3.3	840	6	CA174317 SCROAD101
23	45.6	3.3	914	6	CA106423 SCROAD101
24	45.6	3.3	564	2	BE361762 DGI_82_E0

25	45.4	3.3	660	6	CD231149
26	45.4	3.3	663	6	CB925634
27	45.4	3.3	811	7	CN141741
28	45	3.2	503	8	CC351449
29	45	3.2	824	8	B2628435
30	45	3.2	852	8	CC013120
31	44.8	3.2	568	2	BE775653
32	44.8	3.2	550	2	BE776036
33	44.6	3.2	472	4	BJ210804
34	44.6	3.2	580	4	BJ218083
35	44.6	3.2	604	6	CA253508
36	44.6	3.2	686	6	CA134878
37	44.6	3.2	695	9	CL164460
38	44.6	3.2	942	8	B2575093
39	44.4	3.2	930	9	CL969992
40	44.2	3.2	547	8	BH633034
41	44.2	3.2	672	6	CA194351
42	44.2	3.2	788	9	CC700945
43	44.2	3.2	1483	8	B2573056
44	44.2	3.2	1605	8	B2573144
45	44	3.2	424	9	CL794328

ALIGNMENTS

RESULT 1
LOCUS CL982367 1872 bp DNA linear GSS 21-SEP-2004
DEFINITION ORF5C047480 Oryza sativa Expressed library Oryza sativa (indica
Cultivar-group) genomic, genomic survey sequence.
ACCESSION CL982367
VERSION CL982367.1 GI:5249216
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriophytaceae; Oryzaceae; Oryza.
1 (bases 1 to 1872)
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G.K.S., Deng, X.W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
JOURNAL
COMMENT Contact: Chen Chen
Department of Bioinformatics
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80486676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source
1..1872
location/Qualifiers
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Expressed library"
/notes="Oryza sativa exon trapped genomic sequences"

ORIGIN

Query Match 6.7%; Score 93.4; DB 9; Length 1872;
Best Local Similarity 51.9%; Pred. No. 5.1e-14;
Matches 261; Conservative 0; Mismatches 236; Indels 6; Gaps 2;

OY 633 CGACAGCGCTTGTGTCGTCACCCGCGCGCAATGCGTCAAGCTGAAGCGG 692
DB 552 CGATGACGTCGACGCTGTGTGTCGCGCGCGCGCATGTATGACATGCGCATGG 611
OY 693 CACCATTTTCGTTATTCGTCGCGACGCGTGAACACAGCAGCGCGTTCAGGGCAT 752

Db 612 CAACATGATGAGCTTTTGGCTCGGTGAGCGCGTGAACGACGAGGCGCATTAACGACCGCAT
 QY 753 AGCGGTGCGCGGTATCAACCGATGAAGACTCTGGAAGACAACCGGAAAAAGTCTTCGG
 Db 672 TGGCTTCAACCGCTCCGACGTCGATCTGCGGACATCTGGAATAAAATTCGCG
 QY 813 CGTACCAAAACAATGGGCGGAGAAATACCCCAACCTATCTGGCGGTGACCAAGGCGT
 Db 732 TACGCGTGGGACTGGGTGAGAAAATTCGCAACCGCCGCGCTGGTGAAGCGCT
 QY 873 GATTTCGGCGCGCATCTGCTGACGCGCAATTAACAGAACCGCAAGACCATCA
 Db 792 GATGGAAGCGCGCGGTGATGAGGCGTCCCGGAAAAAAG---TGAACCGCGCA
 QY 933 AATGCTGGCGAAAAACAATACGTGGTCTGACGTGAAAGTCTGGCGGTGACATGAA
 Db 849 GATCTCTTCGCGCGCGCGCTGCTCACTCAAGAAACAGTATCTCAACCGGCGAAATGCT
 QY 993 CGGCACTTTCGAAATACGAAAAAGACATTAACGCGCGCTACCGGACTTTCACACTTCTT
 Db 909 CGGGGAATACGACACGCGCTGGGCGACGCGTGCAGACGCGACCGACCGATCCGTTCTT
 QY 1053 TCGCCACG---CGCAGACTATCCGCTCTTACAGACAGTGCAGTCTGATCTGACCGAGCT
 Db 969 CACGAGAGGTGCGGTGATTAACCTGTAACCTCTCCGACGAGATGTTCTTAACCGAGTT
 QY 1110 GAGCGCTGGGGCATGATCAATG 1132
 Db 1029 CCGCGCTGGGCGCTGCTTAAG 1051

RESULT 2
 A2935061 575 bp DNA linear GSS 24-APR-2001
 LOCUS B1_Ba0003E04r B. japonicum BAC library Bradyrhizobium japonicum
 DEFINITION genomic, genomic survey sequence.

ACCESSION A2935061
 VERSION A2935061.1 GI:13777152
 KEYWORDS GSS.
 SOURCE Bradyrhizobium japonicum
 ORGANISM Bradyrhizobium japonicum
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

REFERENCE 1 (bases 1 to 575)
 Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
 Goltschew,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
 genome

JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
 MEDLINE 21376150
 PUBMED 11483585
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Class: BAC ends
 High quality sequence stop: 535.

FEATURES
 source
 1..575
 Location/Qualifiers
 /organism="Bradyrhizobium japonicum"
 /mol_type="genomic DNA"
 /strain="USD110"
 /db_xref="taxon:375"
 /lab_host="E. coli"
 /clone_lib="B. japonicum BAC library"
 /note="Vector: pindigo536; Site_1: HindIII"

ORIGIN

Query Match 6.3%; Score 87.6; DB 8; Length 575;

Best Local Similarity 61.1%; Pred. No. 1,5e-12;
 Matches 160; Conservative 0; Mismatches 99; Indels 3; Gaps 1;
 QY 131 TCGGCTTCAATCAAGCTTACCGACATGCGCGCGGTGGCGCGCGGAAAAAGCTTCT
 Db 214 TCGGCTTCAATCGGCTGATGATGCGCGCGCGCTGATCTGCGCGCGTGA
 QY 191 TCGAGACGAGGCGCTTCTGTCGAATCTGAAGCGGCAAGCCAACTGAAGGTGATG
 Db 274 CGCGCGCGAAGCGCTGACGTGAACCTGTGCGGAGGCTTCTGTGCAACGTCCGCG
 QY 251 ATAGGCTGTGATGAGCGAAGTGAACGCGCTGCAATGCTGGCGCGCGCGCTTACCGG
 Db 334 ACAAGCTCAATATCGGCTTTCGACGCGCGCATCTGCGCGCGCGCGATGCGCT
 QY 311 CCAGCGTTGGCTTCG---CACCAAGCGCGCATATGAGGTGCGGTTACAGATGGCTTCA
 Db 394 CTTCACTCGGCGCTCGGCGACGTCAAGGTCCGATCGCGCGCGCTTCATCTCGGATCA
 QY 368 ACGGCAACCGCATATACGCTGTC 389
 Db 454 ACGGCAACCGCATACAGGTGTC 475

RESULT 3
 B2575941/c 800 bp DNA linear GSS 17-DEC-2002
 LOCUS msh2_470.x2 msh Pseudomonas aeruginosa genomic clone msh2_470,
 DEFINITION genomic survey sequence.

ACCESSION B2575941
 VERSION B2575941.1 GI:27211002
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

REFERENCE 1 (bases 1 to 800)
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
 Burns,J.L., Kaul,R. and Olsen,M.V.
 Whole-genome-sequence variation among multiple isolates of
 Pseudomonas aeruginosa library
 J. Bacteriol. (2002) In press

JOURNAL Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.

FEATURES
 source
 1..800
 Location/Qualifiers
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="MSH"
 /db_xref="taxon:287"
 /clone_lib="msh2_470"
 /note="Environmental isolate. Whole genomic shotgun
 library."

ORIGIN

Query Match 5.9%; Score 82.4; DB 8; Length 800;
 Best Local Similarity 50.7%; Pred. No. 4.2e-11;
 Matches 225; Conservative 0; Mismatches 216; Indels 3; Gaps 1;

QY 651 GGTACCCCGCGCGCGCAATGCGTCAACGTGGAAGCGGACCAATTTTCGTTATG 710
 Db 526 GGTGTTCCACCGCGCGAGATGTGCGCATCTCCAGGCGCGGCGCATGACGAGGTTCTG 467
 QY 711 CGTGGCGAGCGGTGGAACGACGAGCGGCTTCAAGGCGATAGCGGCGGTATAC 770
 Db 466 CGCGGCGGACCTTGGGCGCGCTGCTGAGACGAGGCGGAGGCTTACCATGCGAC 407

```

QY 771 CGATGAAGAACTCTGGAAGACACCGCGGAAAAAGTCTTGGCGGTGACCAACAAATGGGC 830
DB 406 CACCCAGGCGATCTGGCGCGGACCAACCGGAAAAAGTCTCTGGGCACTACCGCGCTTCGT 347
QY 831 GAGGAATATCCCAACACCTATCTGGGGGTGACCAAGCGCGCTGATTTGGGCGCGGANTCG 890
DB 346 CGAGCGCTACCCCAACACCGCGCGGTGTCTATG3C---GGTCTTGAACGCGCAAGCCG 290
QY 891 GCTGACGCGCGCAATTAACAAGACCGCAAGAACCCATGAAATGCTGGCGGCAAAAACA 950
DB 289 CTTTATCTGAACAGAACCGCCGAAACCGCTGGGACCGCGCACTATGACGCGCGCGCA 230
QY 951 ATATGCTGCTGCTGACGTGGAAGTGTGGCGGCTAGCATGACGCGCACTTTTGAATACGA 1010
DB 229 CTATGTCAGAGCTCCGCTCGCGCGGATCCAGCGCGCTTCTGCGCGCTACACGAGACGG 170
QY 1011 AAAAGACGATTAACGCGCGCTACCGGACTTCAACACTTCTTTCGCCACGCGCGCACTA 1070
DB 169 CCTCGGCAACGCGCTGGGACGACCGCATCCGCTGCTGTATGCGGACGCGCAAGTCAA 110
QY 1071 TCCGCTCTACAGCAGTGCAGTCTG 1094
DB 109 CCGGCGCTGCTCTCGGAGGCTG 86

```

```

RESULT 4
CP161028 406 bp mRNA linear EST 25-JUL-2003
LOCUS B0692F03.5 NIA Mouse Blastocyst cDNA Library (Long) Mus musculus
DEFINITION cDNA clone NIA:B0692F03 IMAGE:30457118 5', mRNA sequence.
ACCESSION CP161028
VERSION CP161028.1 GI:33270577
KEYWORDS Bst.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 406)
Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
JOURNAL MEDLINE 21429098
PUBMED 11544199
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: B0692 row: F column: 03
Seq primer: M13 Reverse
High quality sequence stop: 406
POLYA=No.

```

FEATURES

```

source
1. 406
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="Eukaryota:B0692F03.5"
/db_xref="taxon:10090"
/clone="NIA:B0692F03 IMAGE:30457118"
/issue_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_lib="NIA Mouse Blastocyst cDNA Library (Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were

```

extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-pGATCTGTCTAGATCGCGAGCGCGCCCTTTTCTTTT-3'] from 0.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Loner-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by centrifugation. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-5'. The products were purified by phenol/chloroform and centrifugation. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

Query Match 4.4%; Score 61.8; DB 7; Length 406;
Best Local Similarity 52.5%; Pred. No. 1.5e-05;
Matches 160; Conservative 0; Mismatches 142; Indels 3; Gaps 1;

```

QY 630 CGGCGAGAGCGCTTGTGTGCGTGAACCGCGCGGCAATGCCCTCCACGCTGGAAGC 689
DB 79 CGACACCGATGTCAGATCCGCGCAGTCCGCGCGGCAATGTTGCGAACCTGCGCGC 138
QY 690 CGGACCATTTTGGTATATTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 749
DB 139 CGACACATGACGCGCTTCTCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 198
QY 750 CATAGGCGTGCCTGATGATCAACGATGAAGACTCTGGAAGACACGCGGAAAAAGTCTT 809
DB 199 GGTGCGCTCATCAATCTGACAAAGATATCTGGGAAGGCGATCGTGGCGGT 258
QY 810 CGGCGTACCAACATGCGCGGCGGGAATACCCCAACACTATCTGCGGCGGCGGCGGCGG 869
DB 259 CGGCGCTCCACAGAAATTCCTCAACCGGCGGCAACCTATGCGGCGGCGGCGGCGGCGG 318
QY 870 GCTGAT---TCGGGCGCGGATGCTGCTGCGGCGGCGGCAATATCAAGAACCGCAAGG 926
DB 319 GATCATGATGACGACGCGGCTTCCGCGCAAGGCGGAGACCGCAACGATTCGCGGAGC 378
QY 927 CATCG 931
DB 379 AATCG 383

```

```

RESULT 5
B2548675/c 1085 bp DNA linear GSS 17-DEC-2002
LOCUS pac61-60_1334.g1 pac61-60 Pseudomonas aeruginosa genomic clone
DEFINITION pac61-60_1334, genomic survey sequence.
ACCESSION B2548675
VERSION B2548675.1 GI:27152256
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1085)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In Press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu

```

```

Class: shotgun.
FEATURES             Location/Qualifiers
     source            1..1085
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="1-60"
                        /db_xref="taxon:287"
                        /clone="pacsl-60_1334"
                        /clone_lib="pacsl-60"
                        /note="clinical isolate 1-60 whole genomic shotgun
                        library."
ORIGIN
Query Match          4.2%; Score 58.6; DB 8; Length 1085;
Best Local Similarity 52.2%; Pred. No. 0.00014;
Matches 153; Conservative 0; Mismatches 139; Indels 1; Gaps 1;
QY      GCCGATGTCGGGCGGCGATTTGGCAAGCTGGAAAAGGAAATTTGAAATTCGGCTTCATCA 143
Db       463 GGCCTGGGCGCGCGGCTGTCGAGCGACCGGAAAAGGCCCTCGACATCGGTTTCATGCG 404
QY      144 GCTACCGCAGCATGGCGCGCTGGCGGTGCGCGCCGCAAAAAGGCTTCTTGAGAGCAGGG 203
Db       403 GCTACCGCATTTGGCGCTGCTGATGTGTGGCGCC-ACCATGCTTGGCCAGCCCTACGG 345
QY      204 CCTGTCTGTCGACTGTGAAGCGCGACGCCCAATTGAAAGTGTGTGATGATAGGCTGTGAA 263
Db       344 CCTACGCTCAATCTCAGGCGCGCAACGCTCTGGGCGCACTTGCAGCAAACTGCTCAG 285
QY      264 TGGCAATCTGACGCGCTCGCATGCTGTGGCGCGCGCGCGCTTACGCGCAGCGTTGGCTT 323
Db       284 CGGCGAGCTGATGCGCGCGACAGTGCCTGTACGGGCTGTGTAAGCGCTTCAGCTGGGCTT 225
QY      324 CGGACCAAGGCCGATATACGATGTCGCTTGACATGGGCTTCAACGGCAACG 376
Db       224 CGGCGGCGACGCGCGCAACGAGATGCGGTGTCTATGGAGCTGTGCCAGAACG 172

RESULT 6
LOCUS      BZ563077
DEFINITION BZ563077
ACCSSION   BZ563077
VERSION    BZ563077.1 GI:27186268
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
            1 (bases 1 to 1035)
            Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
            Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
            J. Bacteriol. (2002) In press
            Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source            1..1035
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="2-164"
                        /db_xref="taxon:287"
                        /clone="pacsl-60_1334"
                        /clone_lib="pacsl-60"
                        /note="clinical isolate 2-164 whole genomic shotgun"

```

ORIGIN	Library."
Query Match	3.5%; Score 49.2; DB 8; Length 1035;
Best Local Similarity	55.9%; Pred. No. 0.05;
Matches 113; Conservative	0; Mismatches 88; Indels 1; Gaps 1;
Qy	658 CCGCGCGCGGAAAGCCGTCACAGCTGGAGAGCGGGACACATTTCGGTTATTCGGTGGGC 717
Db	448 CCACCGGGGAGATGTGTGGCCATCTCCAGGCCGGGCGCATCGACAGGGTTCTGGCGCGGC 507
Qy	718 -GAGCCGTGGAACCGACGAGCGCTGTTCAGAGGGCATAGGGCTGCGCGGTATCACCATTGA 776
Db	508 GGAACCTTTGGGGCGCCCTGCGCGGTGAGACGAGGGCAAGGGCTTTCACATCGGCAACAGCA 567
Qy	777 AGAATCTGGAAGACACGCCGGAAAAAGTCTTGCGCGTGAACCAACATGGCGGAGAA 836
Db	568 GGCATTTTGGCGCCGACACACCGGAAAAAGGGCGCTCGAAATCCCGCGCTTCGTTGACGC 627
Qy	837 ATACCCCAACACCTATCTGGCG 858
Db	628 CTACCCGAACACCGCCGCGCG 649
RESULT 7	
BZ572739	
LOCUS	BZ572739 1407 bp DNA linear GSS 17-DEC-2002
DEFINITION	msh2_2775.y2 msh Pseudomonas aeruginosa genomic clone msh2_2775,
ACCESSION	BZ572739
VERSION	BZ572739.1 GI:27207800
KEYWORDS	GSS.
SOURCE	Pseudomonas aeruginosa
ORGANISM	Pseudomonas aeruginosa
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
AUTHORS	Pseudomonadaceae; Pseudomonas.
TITLE	1 (bases 1 to 1407)
JOURNAL	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
COMMENT	Burns,J.L., Kaul,R. and Olsen,M.V.
	Whole-Genome-Sequence variation among multiple isolates of
	Pseudomonas aeruginosa library
	J. Bacteriol. (2002) In press
	Contact: Chris K. Raymond
	Genome Center
	University of Washington
	Box 352145, Seattle, WA 98105-2145, USA
	Tel: 2062216954
	Fax: 2066857244
	Email: craymond@u.washington.edu
	Class: Shotgun.
FEATURES	Location/Qualifiers
source	1..1407
	/organism="Pseudomonas aeruginosa"
	/mol_type="genomic DNA"
	/strain="MSH"
	/db_xref="taxon:287"
	/clone="msh2_2775"
	/clone_id="msh"
	/note="Environmental isolate. Whole genomic shotgun
ORIGIN	library."
Query Match	3.5%; Score 48.6; DB 8; Length 1407;
Best Local Similarity	50.7%; Pred. No. 0.078;
Matches 155; Conservative	0; Mismatches 148; Indels 3; Gaps 2;
Qy	84 GCCCGATGTGCGGCGCACTTGGCAAGCTTGAAAGAAAGATTGAAATTGCGCTTCATCA 143
Db	330 GGCCTGGGCGGCGCGGTTCGACGACCGGAAAGAGCGCCTTGACATCGGTTTCATGGC 389
Qy	144 GCTCACCGACATGCGCGCGCTG-CGGTGGCGCCCGCAAAAAGCTTCTTCAGAGACAGAG 202
Db	390 GCTGACCGATTTGGCTCGCTGATGTGAGCGCGCACACGAGGCTTGGCCAGCGCTTAGC 449

```

Qy 203 GCGTGTTCGTCACCTGGAAAGCGAGCCACTGGAAAGTGATGATAGGCTCTGA 262
    |||
Db 450 GCGTCAAGCTCACTCAGGCGCCAAACCGCTCTGGGCGACCTTGCGGACAACTGCTCA 509
Qy 263 ATGGCGCACTGCGAGCGCTGCGACATGTCGCGCGCGCGGCGTTAGCGCCAGG--TTGG 320
    |||
Db 510 GCGGCGCAGNNGATCGANNNGATCGATGCTGTACGGGCTGTGTTACCGGCGTCCAGCTCGG 569
Qy 321 CTTCGCGACCAAGCGCGATATCGAGGTGCGGCTTCAGCATGGGCTTCAACCGCAACGGCAT 380
    |||
Db 570 CCGCGCGCGAGCGCGCGGCGGAGGCGCGGCTGTCTATGAACTGTGCCAAGACGGCAG 629
Qy 381 TACGGT 386
    |||
Db 630 GCGCAT 635

RESULT 8
BZ550725 788 bp DNA linear GSS 17-DEC-2002
LOCUS paci-60_2795.y2 paci-60 Pseudomonas aeruginosa genomic clone
DEFINITION paci-60_2795, genomic survey sequence.
ACCESSION BZ550725 GI:27154306
VERSION BZ550725.1
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 788)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Haastings,M.,
        Burns,J.L., Kaul,R. and Olsen,M.V.
        Whole-Genome-Sequence variation among multiple isolates of
        Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
        Genome Center
        University of Washington
        Box 352145, Seattle, WA 98105-2145, USA
        Tel: 2062216954
        Fax: 2066857244
        Email: craymond@u.washington.edu
        Class: shotgun.
FEATURES
    source          1..788
                    Location/Qualifiers
                    /organism="Pseudomonas aeruginosa"
                    /mol_type="genomic DNA"
                    /strain="1-60"
                    /db_xref="taxon:287"
                    /clone="paci-60_2795"
                    /clone_1ib="paci-60"
                    /note="clinical isolate 1-60 Whole genomic shotgun
                    library."
ORIGIN
Query Match 3.5%; Score 48.4; DB 8; Length 788;
Best Local Similarity 51.3%; Pred. No. 0.078;
Matches 138; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

Qy 131 TGGGCTTCATCAAGCTCAGCGACATGCGCGCTGCGGTGCGCGCGGCAAAAAGCTTCT 190
    |||
Db 379 TGGCTTACCTGCGATCAGGACGCGCCATGCTGGTGGCCATATGCCAAAGGCGCTGT 320
Qy 191 TCGAAGACGAGGCGCTGTTCTGTCACATGTGAAGCGGCAACTGGAAGTGTGATG 250
    |||
Db 319 TCGAAGACGAGGCGATTCAGCGCGGCGCGGTGCTGTTGCGTACTGCGGCGCAGGTGA 260
Qy 251 ATAGGCTCGTGAATGCGCAATGAGCGGTGCGACATGCTGCGCGCGCGCGCTTAAGCG 310
    |||
Db 259 TCAAGCGCTTCAATCTCCGTCAGGTCAAC---GTCACTCGCGCTGCTGTGCCCATGCCG 203
Qy 311 CCAAGCTTGGCTTGGCAACCAAGCGCGCATATTCAGAGGTGCCGTTCAAGATGGGCTTCAAG 370

```

```

Db 202 TCTGGGCGCGCTACGCGAGCAAGTGTCCGCGCAAGGTGTGGCTTGACACGTCGCG 143
Qy 371 GCAACGGAATTACGCTGTCCATGAATAATC 399
    |||
Db 142 GCTCGGCGCTCAGCGTGTGCTCCAAATGC 114

RESULT 9
CNS0052P 844 bp DNA linear GSS 03-JUN-1999
LOCUS BACR11P6 of RPl-98 library from Drosophila melanogaster (fruit
DEFINITION fly), genomic survey sequence.
ACCESSION AL056652 GI:4932342
VERSION AL056652.1
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 844)
AUTHORS Genoscope.
TITLE Direct Submision
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
        BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr
        - Web : www.genoscope.cns.fr]
COMMENT Determination of this BAC-end sequence was carried out as part of a
        collaboration with the Berkeley Drosophila Genome Project (BDGP).
        The BDGP is constructing a physical map of the Drosophila
        melanogaster genome using these BACs. For further information
        please see http://www.fruitfly.org The BDGP Drosophila
        melanogaster BAC library was prepared by Kazuhiro Oosawa and
        Aaron Mammoser in Pieter de Jong's laboratory in the Department of
        Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
        NY. The library is named RPl-98 and was constructed by partial
        EcoRI digestion of Drosophila DNA provided by the BDGP from the
        isogenic strain y2; cn bw sp, the same strain used for the BDGP's
        P1 and EST libraries. A more detailed description of the library
        and how to order individual BAC clones, the entire library, or
        filters for hybridization from the BACPAC Resource Center can be
        found at http://bacpac.med.buffalo.edu/drosophila\_pac.htm.
FEATURES
    source          1..844
                    Location/Qualifiers
                    /organism="Drosophila melanogaster"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7227"
                    /clone="BACR11P6"
                    /clone_1ib="RPl-98"
                    /note="end : TET3"
ORIGIN
Query Match 3.4%; Score 47.8; DB 9; Length 844;
Best Local Similarity 17.2%; Pred. No. 0.12;
Matches 60; Conservative 143; Mismatches 145; Indels 0; Gaps 0;

Qy 471 AAAACCGTCTCGAATAATCAAAAGCGAAGCCGCTTCAATATGCGCATGACCTT 530
    |||
Db 359 AASASVSAACSSVAGSSMAAGASCBGAANNACGCGSSSSASASRMMSSCSASSS 418
Qy 531 CCCGCGGATTCACAACTCAAACTGCTTACTGCTGCGCGCTGCGCGGTATCAATCC 590
    |||
Db 419 CAGSSSSCGAMSSSCCCTVSSCGMASSSCSGGMSASSSSSGVSSGAVGRCRM 478
Qy 591 TGGCTACTATTGCGCGCGCGAGACATTTCCGGCCAAATGCGCGACAGCCCTTGTTC 650
    |||
Db 479 CCSMVCCCMSCSSMCCSVSSVCAVCSGSVSVAVGCSVGVGRVSSRGRAGRSSGR 538
Qy 651 GGTGACCCCGCGCGCAAAATGCGCCACCGCTGGAAGCGGACCATTTTGGGTATTG 710
    |||
Db 539 GGGSSVSSGVSBSVGMGACSSASVSSCBSSVASSVSSGBVSRGRCRGVGGVGG 598

```


QY	711	CGTGGCAGAGCGGTGGAACACAGACAGCCCGTGTTCAGAGGCATCAAGCGTCGCCGTATGAC	770
Db	599	SRVSCSGSSSSSSGCGSVSRSRSGSANGVRVCGSGRGGCGGGRGCGANNAADDA	658
QY	771	CGATGAAGAACTCTGGAGAAGACACGCCCGGAAAAAGTCTTCGCGCTGAC	818
Db	659	ABRBKALANSNADMGAAVSSVSSVSSBQARRGRDPAALVYNSSVGRB	706

RESULT 10	
CA718866	
LOCUS	CA718866
DEFINITION	CA718866 459 bp mRNA linear EST 26-NOV-2002
ACCSSION	wk2m2n.pk002.c3 wk2m2n Triticum aestivum cDNA clone wk2m2n.pk002.c3 5'
VERSION	CA718866
KEYWORDS	CA718866.1 GI:25440659
SOURCE	EST.
ORGANISM	Triticum aestivum (bread wheat)
	Triticum aestivum

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 459)	Tingley, S. V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z., Mao, G., Caraher, N., and Hanafey, M. K.	Dupont Wheat cDNA Sequence	Unpublished (2002)	Contact: Scott V. Tingley

E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@usa.dupont.com
Seq primer: M13.

FEATURES	SOURCE
Location/Qualifiers	1..459
	/organism="Triticum aestivum"
	/mol_type="mRNA"
	/cultivar="Spring wheat"
	/db_xref="taxon:4565"
	/clone="wkm2n.pk002.c3"
	/tissue_type="kernel"
	/lab_host="DH10B"
	/clone_id="wkm2n"
	/note="vector: pbluescript SK+; Site_1: EcoRI; Site_2: XhoI; wheat (Triticum aestivum L.) kernel malted 175 hours at 4 C, normalized"
ORIGIN	

Query Match	Score	DB	Length
Best Local Similarity	5.4%;	47.2;	459;
Matches	112;	Conservative	0;
		Mismatches	108;
		Indels	0;
		Gaps	0;

RESULT 11
CD894397

LOCUS	CD894397	691 bp	mrna	linear	EST 14-JUL-2003
DEFINITION	G118.126C22F010823 G118	Triticum aestivum	cdna	clone	G118126C22,
ACCESSION	CD894397				
VERSION	CD894397.1	GI:32665884			
KEYWORDS	EST.				
SOURCE	Triticum aestivum (bread wheat)				
ORGANISM	Triticum aestivum				
REFERENCE	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Trilicaceae; Triticum.				
TITLE	1 (bases 1 to 691)				
JOURNAL	Genopiante.				
COMMENT	Genopiante, a major partnership french program in plant genomics Unpublished (2003)				
	Contact: Genopiante				
	Genopiante				
	93, rue Henri Rochefort 91025 EVRY CEDEX France				
	Tel: 33 1 69 47 54 00				
	Fax: 33 1 69 47 54 10				
	This sequence has been generated in the framework of the french plant genomics programme 'Genopiante' (http://www.genopiante.com and http://genopiante-info.infobiogen.fr).				
	location/Qualifiers				
	1..691				

```

/tissue_type="grain (118 degrees per day after
pollination)"
/clone_11b="G1.18"

```

Query Match	Best Local Similarity	3.4%;	Score 47.2;	DB 6;	Length 691;
Matches	112;	Conservative	0;	Mismatches 108;	Indels 0;
					Gaps 0;
QY	156	GCGCCCGCTGCGGGTGGCCGCCGAAAAAGGCTTCTTGGAGGACGAGGCGCTGTTGTCGA	215		
Db	171	GCGCGCATATGGCAGCGCGCAGCAGAGTCGCGACAA CGGAGCCAGTCGGGCCCGACCGGGGGA	230		
QY	216	ACTGGAAAGCCGACGGCCAAACTGGAAAGTGATGATAGGGTCTGTAATGCGCAACTGGA	275		
Db	231	CTCGGGGGCGTTGCTCTTCCTCCCGCGTGAAGCGGAGGAGAGGGCCGCCGAGTCCGAGGCCG	290		
QY	276	CAGCTGCGACATGCTGGCGCCGCGCGCTTAAAGCGGACGCGTTGGCTTCCGACCAAGGC	335		
Db	291	CGCTTCGCGCGCGCAGCGCGTCCGAGTCTCGAGAACGAGGGTCCAGCAGCGCAACATCCA	350		
QY	336	CGATATCGAGGTGCGCTTCAAGCATGGGCTTCAACGGGAC	375		
Db	351	GGAAGCCGAGTCTCTCGCTTCGCGAGGGGCTTCTCTTCAAC	390		

RESULT 12	
CN129290	
LOCUS	CN129290 824 bp mRNA linear EST 01-APR-2004
DEFINITION	RHOH_34_C09.g1_A002 Acid- and alkaline-treated roots Sorghum
ACCESSION	bicolor cDNA clone RHOH_34_C09_A002 5', mRNA sequence.
VERSION	CN129290
KEYWORDS	CN129290.1 GI:45956911
SOURCE	EST.
ORGANISM	Sorghum bicolor (sorghum)
	Sorghum bicolor
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
	clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE	1 (bases 1 to 824)
AUTHORS	Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,
	Sun,F., Sullivan,R., Lim,S., Eastman,A. and Pratt,L.H.
TITLE	An EST database from Sorghum: acid- and alkaline-treated roots

**JOURNAL
COMMENT**

Unpublished (2003)
Other_ESTRs: RH01H_34 C09.b3_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210

FEATURES

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sugs (CTTCTGCTCTAAAGTCGG) .

Location/Qualifiers

```

/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4556"
/clone="KH01_34_C09_A002"
/lab_host="DH08-T1 phage-resistant E. coli"
/clone_1db="Acid- and alkaline-treated roots"
/notice="Organ: Root; Vector: pME185-PL3, Site: 1; XhoI; Site 2; XhoI; The library was prepared from pLY4+ RNA from 8-day-old roots harvested from BTx623 sorghum seedlings grown in hydroponic culture. HCl was added to a pH of 3.0 to some seedlings, KOH to a pH of 9.0 for others."
Roots were harvested 3, 12 and 27 hr after addition and pooled for RNA preparation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME185-FL3 vector (5-prime DraIII site is CACGATGG, 3-prime DraIII site is CACCATGG). XhoI excisees the cDNA insert."

```

ORIGIN

Query Match	3.48;	Score 47;	DB 7;	Length 824;
Best Local Similarity	52.3%;	Pred. No. 0.19;		
Matches 104;	Conservative 0;	Mismatches 95;	Indels 0;	Gaps 0;

QY	DB	QY	DB	QY	DB	QY	DB
156	GGCGCTGGCGGCTGGCGCGCGAAAAGGCTTCTTGAGGACGAGAGGGCGCTTCTCGGCACT	218		219	GGAAAGCGCAGCGCAACTGGAGAGTGTGATGTGATGAGTCTGTGATGTGGCACTGGACGG	278	
319	GCTGCTGCGGCTCTACTTCTTCCACGACTGCTTCCCGCAGGGCTGGCAGCGCTGGTGGCT	378		379	GGACGGCGCGGGGCACGAGAGAGGGCGCGCGCGCAACAGTCTGCTGGGGCGGGCTGGACGT	438	
439	CTTCGACATGCTGGCGCGCGCGCGCTTGAAGCGGCAGCGTTGCGCTTCCGACCAAGGCGCA	498		499	CGTCGACGCGCGCAAGCGCGCGCTGGAAGCGCGCTGCGGGCACCGCTCTTCGCGGGA	517	
517	CGTCGACGCGCGCGCGCGCGCGCGCTTCAAGC	576					

RESULT 13

LOCUS	623 bp	mRNA	linear	EST 26-NOV-2003
DEFINITION	wlk8.p00022.g9 wlk8	Triticum aestivum	cDNA clone	wlk8.p00022.g9 5'
ACCESSION	CA6996640	end, mRNA	sequence.	
VERSION	CA6996640			
KEYWORDS	CA6996640.1	GI:25421433		
SOURCE	EST.			
ORGANISM	Triticum aestivum (bread wheat)			
	Triticum aestivum			
	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			

REFERENCE

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Pooledee; Triticasee; Triticum.
1 (bases 1 to 623)
Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,
Miao, G., Caraher, N. and Hanfey, M.K.
Duront Wheat cDNA Sequence
Unpublished (2002)
Contact: Scott V. Tingey

FEATURES

```

/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Stephens"
/db_xref="taxon:4565"
/clone="w1k8.pk0022.g9"
/tissue_type="leaf"
/clone_lib="w1k8"
/note="Vector: pGEM-7; Site_1: EcoRI; Site_2: XhoI; wheat
Triticum aestivum L.) seedlings 8 hr after treatment with
6-iodo-3-propyl-2-propyloxy-4(3H)-quinolinone"
```

ORIGIN

Query Match	3.3%;	Score 46;	DB 6;	Length 623;
Best Local Similarity	46.3%;	Pred. No. 0.33;		
Matches 136; Conservative	0;	Mismatches 158;	Indels 0;	Gaps 0;

Qy	158	CGCCGCTGGGGTGGTGGCCCGCGAAGAAAGCTTCTTCCAGGACGAGGGCGCTGTGTGGAC	217
Db	51	CGAGGTGGTGGTGGCGCGCGAGAGGGGTTCAGCGAGATGCGCGCTGGGGCCAGAGG	110
Qy	218	TGAGAGCGCAGCGCAACTGGAGGTGTGTGATAGGTGCTGTGAATGCGCAACTGACG	277
Db	111	CGAGGCGAATGATCCCGGGCGGTGTGGCAGCGACAGAGGGGGGTCTNGATGATGTGTCATGGG	170
Qy	278	GCTTCGCATGCTTGGGCGCGCGCCGCTTACGGCGCAGCGCTTGGCTTGGCACCAAGGCG	337
Db	171	GCGAGGAGGCGAGCGGATACCCCGCGCGCGCGAGGGGTCTCCAGAGGCTCAAGAAC	230
Qy	338	ATATCAGGTCGCGTTACGATGGGCTTCAACGGCAACGGATTCAGGTGTCCATGAAA	397
Db	231	GAGGTGATGACGTGTGTCTCTCGTGTCAAAAGCTAAATTGAAGCGCGGACGGTG	290
Qy	398	TCTGGCATCAGATGAAAGCGAATACCGCTGGAGGCGGTAAACGGTGATC	451
Db	291	TACCGTCCAGAGGAAGAGAAAGACATCTTCGCGGTTTCGGCGCTGCGTC	344

RESULT 14

LOCUS	469 bp	mrna	EST 23-SEP-2003
DEFINITION	linear	clone	SC6QAD1015H11
ACCSSION	SC6QAD1015H11	g	AD1 Saccharum officinarum
ORIGIN	5' mRNA sequence.		

ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 469)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
The libraries that made SUCST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2005, 17:32:09 ; Search time 171 Seconds
(without alignments)
1049.455 Million cell updates/sec

Title: US-10-689-200-2

Perfect score: 2443
Sequence: 1 MTTIRSSSKLLTLTSLA.....AKPAIGKQTVAGKVD 464

Scoring table: BLAST62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2000s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2443	100.0	464	8	AD006543 Methylo
2	577.5	23.6	456	7	AB063846 Klebsiell
3	525	21.5	403	6	ABU42051 Protein e
4	446.5	18.3	474	8	ADL05206 M. catarr
5	338.5	13.9	336	6	ADA36285 Acinetoba
6	292.5	12.0	249	7	AB081755 Pseudomon
7	240	9.8	399	8	ABU38446 Protein e
8	240	9.8	399	8	ADSL1947 Pseudomon
9	240	9.8	419	7	AB076961 Pseudomon
10	234.5	9.6	396	6	ABU21883 Protein e
11	227	9.3	257	6	ABU16603 Protein e
12	226	9.3	294	4	AA091152 C glutamyl
13	222.5	9.1	372	6	ABU40439 Protein e
14	220.5	9.0	391	6	ABU22033 Protein e
15	219.5	9.0	275	4	AA076701 Corynebact
16	217.5	8.9	391	6	ABU02558 Protein e
17	157	6.4	197	7	ABO81646 Pseudomon
18	144	5.9	342	6	ABU44679 Protein e
19	140	5.7	192	6	ABU19752 Protein e
20	133.5	5.5	319	7	ADM25995 Hyperther
21	118.5	4.9	939	4	AB071135 Pseudomon
22	116.5	4.8	337	7	ABO78011 Pseudomon
23	116.5	4.8	623	7	ABO77064 Pseudomon
24	113	4.6	1355	4	ABG05726 Novel hum
25	111	4.5	1357	4	ABG09635 Novel hum

26	106.5	4.4	461	6	ABU18735 Protein e
27	106	4.3	433	5	ABBS4838 Lactococc
28	106	4.3	933	8	ADQ66768 Novel hum
29	106	4.3	1190	4	AA078353 Human pro
30	105	4.3	419	6	ABU18513 Protein e
31	105	4.3	1190	5	AAU74335 Human cyt
32	105	4.3	4455	7	ADJ95076 Novel NOV
33	104	4.3	866	8	ADL90386 Clostridi
34	104	4.3	1420	8	ADL90392 Clostridi
35	103.5	4.2	280	4	AAU36121 Klebsiell
36	103.5	4.2	337	7	ABO65619 Klebsiell
37	103	4.2	864	8	ADL90384 Clostridi
38	102.5	4.2	320	6	ABU27850 Protein e
39	102	4.2	369	7	ADC96127 E. faeciu
40	102	4.2	613	8	ADS29206 Bacteriat
41	102	4.2	860	8	ADL90457 Clostridi
42	102	4.2	862	8	ADL90376 Clostridi
43	102	4.2	866	8	ADL90370 Clostridi
44	102	4.2	867	8	ADL90380 Clostridi
45	102	4.2	867	8	ADL90378 Clostridi

ALIGNMENTS

RESULT 1
AD006543
ID AD006543 standard; protein; 464 AA.

AC AD006543;
DT 29-JUL-2004 (first entry)
XX

.Methylomonas sp 16a nitrogen transporter ntrA protein.

XX C1 metabolizing bacteria; gene expression regulation;
KM nitrogen transporter; ntrA.
KW

OS Methylomonas sp.; 16a.
XX

PN W02004037998-A2.
XX

PD 06-MAY-2004.
XX

PF 21-OCT-2003; 2003W0-US033698.
XX

PR 21-OCT-2002; 2002US-0419872P.
XX

PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX

PI Dicosimo DJ, Picataggio SK, Selp JE, Ye RW, Wang T, Ni H;
XX

DR WPI; 2004-375911/35.
XX

DR N-PSDB; AD006542.
XX

PT Promoter regions responsive to presence of nitrate, acidic pH, elevated temperatures or highly expressed in presence of methane or methanol
PT useful for driving expression (especially for producing zeaxanthin) in C1
PT metabolizing bacteria.
XX

PS Claim 18; Page 70-71; 83pp; English.
XX

XX The present invention relates to a method of expressing a coding region
XX of interest in a C1 metabolizing bacteria. Also provided are protein
XX coding and promoter sequences from Methylomonas sp. 16a which respond
XX to various metabolic and growth conditions, which can be used in the method
XX of the invention. The promoters are useful for driving expression of a
XX number of coding regions (especially for producing zeaxanthin) in C1
XX metabolizing bacteria under controlled conditions. The present sequence
XX is a Methylomonas sp. protein.
XX

SQ Sequence 464 AA;

Query Match	100.0%	Score 2443	DB 8	Length 464
Best Local Similarity	100.0%	Pred. No. 1e-217		
Matches 464	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MKTIIRSSSKLLLTLSASLAWGTLTIPDVGAVKLEKEDIKREFITKLTQMAPLAAVAE	60	
Db	1	MKTIIRSSSKLLLTLSASLAWGTLTIPDVGAVKLEKEDIKREFITKLTQMAPLAAVAE	60	
QY	61	KGFPEDEBLFYOLEAQAAMKVMDERVNAGELDGSHMLAPAPLAAVGFYTKADIEVPSM	120	
Db	61	KGFPEDEBLFYOLEAQAAMKVMDERVNAGELDGSHMLAPAPLAAVGFYTKADIEVPSM	120	
QY	121	GFNGNAITVSNIEIHQMKPNIPLEGKRVHPKADYIKRVYEKYABGKPFNMAMTFPAG	180	
Db	121	GFNGNAITVSNIEIHQMKPNIPLEGKRVHPKADYIKRVYEKYABGKPFNMAMTFPAG	180	
QY	181	SHNFKLRFWLAAAGINPGYSPPODISGQIGADLLSTTPPQMSLTIEAGTIFGYCIGE	240	
Db	181	SHNFKLRFWLAAAGINPGYSPPODISGQIGADLLSTTPPQMSLTIEAGTIFGYCIGE	240	
QY	241	PMNOQAVFKGIGVPYITDEELMKDPEKVFQVTKQMAEKYPNTVLAATKALIRAIWIDA	300	
Db	241	PMNOQAVFKGIGVPYITDEELMKDPEKVFQVTKQMAEKYPNTVLAATKALIRAIWIDA	300	
QY	301	DNNKRRKEAIEMLAQKQYVGADVEVLAASMGTFEYEKDQRALPDENTFFRHGASYPY	360	
Db	301	DNNKRRKEAIEMLAQKQYVGADVEVLAASMGTFEYEKDQRALPDENTFFRHGASYPY	360	
QY	361	SSATWYLLTQLRRKMGINFEKPDNWLDTAKRVYRPDIYLAALKEIVAGKAKAEDFPADT	420	
Db	361	SSATWYLLTQLRRKMGINFEKPDNWLDTAKRVYRPDIYLAALKEIVAGKAKAEDFPADT	420	
QY	421	SIKSQNFIDKVPEDANKPNDYIAKFAIGLKQTVAGGKVV	464	
Db	421	SIKSQNFIDKVPEDANKPNDYIAKFAIGLKQTVAGGKVV	464	
RESULT 2				
ABO63846				
ID	ABO63846	standard; protein; 456 AA.		
AC	ABO63846;			
XX				
DT	29-JUL-2004	(first entry)		
DE				
KM	Klebsiella pneumoniae polypeptide segid 10363.			
KW	Recombinant expression vector; transcription regulatory element;			
XX	Klebsiella pneumoniae protein; antibacterial; vaccine.			
OS	Klebsiella pneumoniae.			
XX				
PN	US610836-B1.			
PD	26-AUG-2003.			
PF	27-JAN-2000; 2000US-00489039.			
XX				
PR	29-JAN-1999; 99US-0117747P.			
PA	(GENO-) GENOME THERAPEUTICS CORP.			
PI	Bretton GL; Osborne M;			
DR	WPI; 2003-895346/82.			
DR	N-PSDB; ACH97397.			
PT	New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for			
PS	preparing a vaccine composition against Klebsiella pneumoniae.			
XX				
XX	Disclosure; SEQ ID NO 10363; 932PD; English.			
XX				
CC	The invention describes a new isolated nucleic acid encoding a Klebsiella			

[illegible]

XI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zykkind JW;
 X0 Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 X9 WPI; 2003-029926/02.
 X8 N-PSDB; ACN455921.
 X7
 X6 New antisense nucleic acids, useful for identifying proteins or screening
 X5 PT for homologous nucleic acids required for cellular proliferation to
 X4 PT isolate candidate molecules for rational drug discovery programs.
 X3
 X2 Claim 25; SEQ ID NO 69975; 1766bp; English.
 X1
 X0 The invention relates to an isolated nucleic acid comprising any one of
 the 6213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required
 for cellular proliferation to isolate candidate molecules for rational
 drug discovery programs, or for screening homologous nucleic acids
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
 the target prokaryotic essential genes. Note: The sequence data for this
 patent did not form part of the printed specification, but was obtained
 in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pat_sequences
 Sequence 403 AA;

[illegible]

Db 288 SDELGNQWQDDPHAVSEFHQCGQVNYVFWLSDGMFWMTQFPRWGLARE-DPD-VYLAVASRVQ 344

QY 394 RPDYIYLAAKELVABGKAKALEDPPADTSIRKSPNFFIDKVPFPAANKENDYLAKEAI 449

Dd 345 QLDLRYQAANAL-----GIDAPSAIT-LRSSQ--LIDGKWKDSDSDAGVARSKTL 390

XX	RESULT 4
XX	ADL05206
ID	ADL05206 standard; protein; 474 AA.
XX	
XX	ADL05206;
AC	
DT	06-MAY-2004 (first entry)
XX	
DE	M. catarrhalis protein #972.
XX	
KW	Moraxella catarrhalis; infection.
XX	
OS	Moraxella catarrhalis.
XX	
PN	US6673910-B1.
XX	
PD	06-JAN-2004.
XX	
PP	04-APR-2000; 2000US-00540236.
XX	
PR	08-APR-1999; 99US-0128416P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Breton GL;
XX	
DR	WPI; 2004-178127/17.
XX	
DR	N-PSDB; ADL03286.
XX	
PT	New nucleic acid encoding a Moraxella catarrhalis polypeptide, useful for
PT	preparing a composition for diagnosing, preventing or treating infection
XX	caused by Moraxella catarrhalis.
XX	
PS	Disclosure; SEQ ID NO 2892; 429pp; English.
CC	
CC	The invention relates to an isolated nucleic acid encoding an Moraxella
CC	catarrhalis polypeptide. The nucleic acid is useful for preparing a
CC	composition for diagnosing, preventing or treating infection caused by
CC	Moraxella catarrhalis. The present sequence represents the amino acid
CC	sequence of a M. catarrhalis protein.
XX	
XX	Sequence 474 AA;
XX	

	Query Match	18.3%	Score 446.5; DB 8;	Length 474;
	Best Local Similarity	27.9%;	Pred. 3.4e-32;	
	Matches 129;	Conservative 82;	Mismatches 160;	Indels 71; Gaps 11;
Oy	2 KTIIRSSSKLLLTLSLASLAWGLTIAPDVAGVKL--EKEDLKRGFKLTDMAPLAVAA	59		
Dd	79 KTVKXTKA-----IAAIASYLPITATLOEAAAIDTLRKPKASVIGFPIICAPPITMAD	132		
Oy	60 EKGFEEDEGLEVOLEAOANMKVWMDRVVNGELDSHMLAPPLASVGFGT-KADIIEVPF	118		
Dd	133 PLGYTAAGGIKANILTKRAQNALVDDQMNRRLDAAHLPAHPPIAININGLSAKONKVAA	192		
Oy	119 SMGFNNAITVSNEIHMOKKNPILDGGKPVHPIKADYLKEVEEIKYAEGKFENAMTFP	178		
Dd	193 IONTNGALVVA-----LKHONRNRPKNMGMTFAIPFE	226		
Oy	179 AGSHNIKLRWYLAAGINPCYYSPPODISQIGADALLSVTPPEMESTTEAGTIFGCYCV	238		
Dd	227 HSIINYLLRFFLAHBGIDP-----DDVDVLRILTTPPDMAIANLKAGNIIDGEPFG	273		
Oy	239 GEPPNQOAVEKGIVGVYITDELKMDTPBEKKVFYTTKOAEKYTPNTTYLATVALRAAIWL	298		
Dd	274 PEPPNQAAMDKAGYIHTLSRDINDNGHPCSGFTGSOFINDYPOTFLAMYAFIIKANVM-	332		

acne; periodontal disease.
 Pseudomonas aeruginosa.
 WO2004083385-A2.
 30-SEP-2004.
 11-MAR-2004; 2004WO-US007467.
 14-MAR-2003; 2003US-00389647.
 (IOWA) UNIV IOWA RES FOUND.
 Greenberg EP, Schuster M, Lostroh C,
 WPI; 2004-709932/69.
 Identifying a modulator of quorum sensing signaling in bacteria, useful
 for treating a biofilm-associated disorder, comprises contacting the cell
 with a quorum sensing signal molecule in the presence and absence of a
 test compound.
 Disclosure; SEQ ID NO 502; 233pp; English.

The present invention relates to a method for identifying a modulator of
 quorum sensing signaling in bacteria. The method comprises: providing a
 cell that comprises a quorum sensing controlled gene (ADSI4446-ADSI4798),
 where the cell is responsive to a quorum sensing signal molecule such
 that a detectable signal is generated; contacting the cell with a quorum
 sensing signal molecule in the presence and absence of a test compound;
 and detecting a change in the detectable signal. The method and modulator
 identified by the method are useful for treating a biofilm-associated
 disease or disorder, e.g. cystic fibrosis, AIDS, middle ear infections,
 acne, periodontal disease, catheter-associated infections, and medical
 device-associated infections. The present sequence is a quorum sensing
 controlled protein, used to illustrate the invention. Note: The sequence
 data for this patent was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences

Sequence 399 AA;

Query Match 9.8%; Score 240; DB 8; Length 399;
 Best Local Similarity 24.6%; Pred. No. 4,1e-13;
 Matches 93; Conservative 68; Mismatches 141; Indels 76; Gaps 14;

```

9 SKKLLITTSASLAWG-LTIAPDVGAVGKLEKE-DIKFGFTKLTDMAPLVAAEKGFED 66
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
9 SRDILKLAALISAAGALPLLSLQARAAABPDAPVRIGYLPITDPTLVVHANGLFEA 68
67 EGLFVQLEAQA--NMKVVMMDRVNCGELDSHMLAPAPLAASVGCTKADIEVPSMGFN 123
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
69 EG--IOAEPPVILRSWAQVTEAFISQGVNVVHLSPMTWAR--YGSKPAPVAVAMNHVG 124
124 GNAITVSNIEIWMQKNPIPLEGSKPVHPKADYLPKVEKYEKGFPMAMTFPAGSHN 183
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
125 GSGLTIVAPEI-----ADVROLGGSKV-----AIPFWYSIHN 155
184 IKLRVWLAAGINPGYSPDIDISQIGA-----DALLSVTPPPQMPSTLEAGTIFGYCV 238
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
156 VVLQQLLRNGI-----RAVSRVAGAAALADENVLVLPSPDMPPALASKRIHGIV 207
239 GEPNMQOAVFKGIGPVITDEELMKDTPKRVGVTKOMAEKYPNTYLAATKALIRAAIVL 298
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
208 ABPPALAKENLVKRGQRFQFTGDVWRNHACCVVFMHEHLERRPQMSQKVVAIVKQQLM- 266
299 DADNNKIKREATEMLAQ---KQYVGADYVLT-----AASMGTFPEYKDDKKA 343
267 ---TEHREAAQOLSKAGANRYTPHAPVIGRVILPAGAEEOAYILASGIRHADMQRER 323
344 LPDENTFFRHGASPSYS 361
324 I-DFOPY-----PYPSYT 335

```

RESULT 9
 AB076961
 ID AB076961 standard; protein; 419 AA.
 XX

AB076961;

29-JUL-2004 (first entry)

Pseudomonas aeruginosa polypeptide #9136.

Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

Pseudomonas aeruginosa.

US6551795-B1.

22-APR-2003.

18-FEB-1999; 99US-00252991.

18-FEB-1998; 98US-0074788P.

27-JUL-1998; 98US-0094190P.

(GENO-) GENOME THERAPEUTICS CORP.

Rubenfield MJ, Nolling J, Deloughery C, Bush D;

WPI; 2003-615309/58.

N-PSDB; ABD10532.

Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
 useful as molecular targets for diagnostics, prophylaxis and treatment of
 pathological conditions resulting from bacterial infection.

Disclosure; SEQ ID NO 25707; 455pp; English.

The invention relates to Pseudomonas aeruginosa polypeptides and the
 CC polynucleotides encoding them. The sequences are useful in diagnosis and
 CC therapy of pathological conditions, as molecular targets for diagnostics,
 CC prophylaxis and treatment of pathological conditions resulting from a
 CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-P. aeruginosa drugs, as templates for recombinant
 CC production of P. aeruginosa-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of P. aeruginosa-caused
 CC infection, and in detection of P. aeruginosa sequences or other sequences
 CC of Pseudomonas species using biochip technology. Sequences AB067826-
 CC AB084396 represent P. aeruginosa polypeptides of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC segdata.uspto.gov/sequence.html

Sequence 419 AA;

Query Match 9.8%; Score 240; DB 7; Length 419;
 Best Local Similarity 24.6%; Pred. No. 4,4e-13;
 Matches 93; Conservative 68; Mismatches 141; Indels 76; Gaps 14;

```

9 SKKLLITTSASLAWG-LTIAPDVGAVGKLEKE-DIKFGFTKLTDMAPLVAAEKGFED 66
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
29 SRDILKLAALISAAGALPLLSLQARAAABPDAPVRIGYLPITDPTLVVHANGLFEA 88
67 EGLFVQLEAQA--NMKVVMMDRVNCGELDSHMLAPAPLAASVGCTKADIEVPSMGFN 123
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
89 EG--IOAEPPVILRSWAQVTEAFISQGVNVVHLSPMTWAR--YGSKPAPVAVAMNHVG 144
124 GNAITVSNIEIWMQKNPIPLEGSKPVHPKADYLPKVEKYEKGFPMAMTFPAGSHN 183
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
145 GSGLTIVAPEI-----ADVROLGGSKV-----AIPFWYSIHN 175

```

QY 184 IKRLVLAAGINPGYSPDISGQIG-----DALSTPPPMPTLEAGTIRGCV 238
 DB 176 VVLAQQLLRNGL-----RAVSRVAGALAADEVNLVYLPPSDMPALAKRIRHGYIV 227
 QY 239 GEPWNOQAVFGKIGVPIITDEELMKDPEKVGVTWKOMAEKYRNTYLAVTKALIRAAIWL 298
 DB 228 AEFNALAEUKKRGQRFGTGDVWRNACCVFMHEHDLERRRQMSKVNAIVKQOLM- 286
 QY 299 DADNNKRRKAEIEMLAQ---KQYVADVEYL-----AASNNGTFEYEKDKRA 343
 DB 287 ---TRHRAEAQAQLLSKAGANRYTPHAPVTLGRVLAAPGAEQQAAYLASGAIIRADQERR 343
 QY 344 LPDPFTFRRGASPSYS 361
 DB 344 I-DFOPY-----PYPSYT 355

RESULT 10

ABU21863
 ID ABU21863 standard; protein; 396 AA.

AC ABU21863;

XX 19-JUN-2003 (first entry)

XX Protein encoded by prokaryotic essential gene #7410.

DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Burkholderia fungorum.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002MO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELITR-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Treweek JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX MPI; 2003-029926/02.

XX N-PSDB; ACA25753.

XX Claim 25; SEQ ID NO 49807; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX proliferation or the activity of a gene in an operon required for

XX proliferation; (7) identifying a compound that influences the activity of

XX the gene product or that has an activity against a biological pathway

XX required for proliferation, or that inhibits cellular proliferation; (8)

XX identifying a gene required for cellular proliferation or the biological

XX pathway in which a proliferation-required gene or its gene product lies

XX or a gene in which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 396 AA:

Query Match 9.6%; Score 234.5; DB 6; Length 396;
 Best Local Similarity 23.3%; Pred. No. 1.3e-12;
 Matches 105; Conservative 77; Mismatches 150; Indels 119; Gaps 18;

QY 9 SKKLTLTSLASLAWGTLTAPDVGAVKLEKED-----LKRFGIKLDMAPLAAAEKGF 64
 DB 7 SRREWLKLSMFTVAG--AAPLIALNARAENRDPVRIGYLPITDAAPLVAHNNGYF 64
 QY 65 EDEGLFVQLEA--QANKVVDRAVYNGELDSHMLAPAPLAASVFGTKADIEVPFSGFN 123
 DB 65 DASGLAVEKPTLLRSNAQLVEAFLSQVNVVHILAPMTIAR--YSGQAPAKVVAHNVN 122
 QY 124 GNAITVSNEIWHQKPNIPLEGKPVNPIKADYLKPEVVEKYAKBGRFNNAMTFPAGSHN 183
 DB 123 GSALTVAPOI-----GKLGEIGGKTV-----AVPFWSIHN 153
 QY 184 IKRLVLAAGINPGYSPDISGQIGDAL--LSVTPPOMSTLEAGTIRGCVGSEPW 242
 DB 154 VVVQHMILRAQGL-----VPLERKDELKANEVLLVWSPDMPALASROIAGIYAEPP 208
 QY 243 NQOAVFGKIGVPIITDEELMKD-----TRKVFQVTWKOMAEKYRNTYLAVTKALIRAA 295
 DB 209 NAAAEELKVGKVLRFQGDVWKNACCVFMHEHDLERBAWSQK-----VVDAAVKAQ 261
 QY 296 IWLADADNNKRRKAEIEMLA---OKQYVADVEYLAASWNGTFEYEKDKKALPDFTFFR 352
 DB 262 VMTFRA-----HPQAAQLLSKSGRNHYTPHSANVLT---NVLAPPGDEGRYLD----- 308
 QY 353 HGASYSYSAAVWYVTLQLRKMGWINEFKPDNWLIDRAKNV--YRPDIYLAAXELV----- 406
 DB 309 -----RAITHADWH-----AKRIDFQPYPPAYATEELVRLKA 341
 QY 407 --AEGKAK-----AEDFPADTSIKPS 425
 DB 342 TQVEGNNQPLHLDPAFVARDLVDRFKVS 372

RESULT 11

ABU16603
 ID ABU16603 standard; protein; 257 AA.

AC ABU16603;

XX 19-JUN-2003 (first entry)

XX Protein encoded by prokaryotic essential gene #2130.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX *Acinetobacter baumannii*.

XX WO200277183-A2.

XX 03-OCT-2002.

XX

PF 21-MAR-2002; 2002MO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;
 PI Wall D, Trawick J, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
 XX
 DR N-PSDB; ACRA0473.
 DR WPI: 2003-029926/02.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 44527; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-regulated gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed, (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SO Sequence 257 AA;

Query Match 9.3%; Score 227; DB 6; Length 257;
 Best Local Similarity 23.5%; Pred. No. 3.4e-12;
 Matches 69; Conservative 52; Mismatches 122; Indels 50; Gaps 7;

QY 91 LFGSHMLAPAPLAASVGFRTKAD-----TEVPSMFPNGNATIVSSEIWHQM-----KPMI 141
 DB 1 LPAAHGXSMPLPAAANG-----ADQIGIALQTPPLVLSKNNAPFSLSSQKLIYQLAIOESDVA 56
 QY 142 PLEGGKPVHPIKADYLVKPVVEKYEKGEKFPNNMAMTPPAGSHNIKLYWTLAAGINPGYS 201
 DB 57 QTTAAQVIVYIGD-----TUSLAHVFKSHIHXYCAREWALA----- 95
 QY 202 PPQDISGQIGADALLSVTPPQPMSTLEAGTIFGCVGSEPMNQAVFKSIGVPVITDEEL 261
 DB 96 -----DSRIAQTLTKIKALPPPYVVEALDHNVIDGFCVGEPPMNTQSGELIGLSKIVCSSQDI 150
 QY 262 WKDTPEKVGSGVKQMAEKYPTNYLATATKALIRAIWLDADNNKRNKREALEMLAQKYVGA 321

DB 151 IPVNVADKVALAVTQEMAEQHPQTLVALTATIMKAO--KEISNLEADPAPILKTL----- 200
 QY 322 DVEVLAASWNGTFEVEKDKRALPDNTEFRHGSYPSYSSAWYLTOLRRWG 374
 DB 201 -VEFGIVRFHCSBEVAVDKRYMIGNIVKLVENANAPQEDDFHMLRQGMQKMG 252

RESULT 12

AAAG91152
 ID AAAG91152 standard; protein; 294 AA.

XX
 AC AAAG91152;

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 4906.

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

OS Corynebacterium glutamicum.

PN EPI108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-00127688.

PR 16-DEC-1999; 99QP-00377484.

PR 07-APR-2000; 2000JP-00159162.

PR 03-APR-2000; 2000JP-00280988.

XX (KYO) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR N-PSDB; AAH66371.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analyzing

PT expression profile or pattern of a gene and identifying homologous gene.

XX Claim 17; SEQ ID NO 4906; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of Corynebacterium bacteria, measuring expression amount and analyzing

CC the expression profile or expression pattern of a gene derived from

CC Corynebacterium bacteria, and identifying a homologue of a gene derived from

CC Corynebacterium bacteria. Corynebacterium bacteria are useful for producing amino

CC acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described in the

CC exemplification of the invention. Note: The sequence data for this patent

CC did not form part of the printed specification, but was obtained in

CC electronic format directly from the European Patent Office

SO Sequence 294 AA;

Query Match 9.3%; Score 226; DB 4; Length 294;
 Best Local Similarity 23.6%; Pred. No. 5.2e-12;
 Matches 78; Conservative 56; Mismatches 152; Indels 44; Gaps 6;

QY 1 MTTIRSSSKLLTLTSLASLAWGLTIPDVQAVKLEKED--LKFGRIKLTDMAPLAV 57
 DB 1 MTHHIFDSRRFQIQAFSLSTPLAGAAKYVSTSNNEPADTPTLTIGVPVIGASAPYAI 60
 QY 58 AAKEGFEEDEGFEVQLEQANKVWMDRVNNGELDGSHTLAPAPLAASVGF-GTADYIV 116
 DB 61 ADALGLFKAGVNVTLKKTSGHSDIMTATVATQQLDVAHMLSPWYAINAGVTNARPTL 120

QY 117 PFSMGFNGNAITVSNELIWHQMKPNIPLEGKPVHPRIKADYLKPVVEKYKAEKGFNNAMT 176
 Db 121 SFTONTGQAITLASKHYGSVNSADLK-----MVLGIFPEYSV- 160
 QY 177 PFMGSHIKIRYMLAAGINPGYISPPQDISGQIGADALISVTPPPMPSTLENGTIFGY 236
 Db 161 -----HALLRDYLVNSAVDP-----IADLERLRLPADMVNQLTYEGIDGF 202
 QY 237 CVGEPNMQAVFKIGVPIVITDEELMKOTPEKVGTVKQNAEKYENTYLAVTKALIRAAI 296
 Db 203 ICGEPFERAIRISNSGRIMWLTQKMDKPCCAVAMAKEMKAPTAAGVLAALREASA 262
 QY 297 WLDADNNKRRKAEIMLAQKQYGVADYEV 326
 Db 263 IL--SNPAGFDS\$ARTLSQEKYLNQPATLL 290

RESULT 13

ABU0439 ID ABU0439 standard; protein; 372 AA.

XX AC ABU0439;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #25966.

XX KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Pseudomonas putida.

XX PN MO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948893.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.

XX DR N-PSDB; ACA44309.

XX PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 68363; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 372 AA;

Query Match 9.1%; Score 222.5; DB 6; Length 372;

Best Local Similarity 23.1%; Pred. No. 1.6e-11;

Matches 90; Conservative 71; Mismatches 139; Indels 89; Gaps 17;

QY 42 LKGFYKLTDMAPLAVAEKGFEDGFLVQLER-QANMKVVDVNVNGLDSHMLAPA 100
 Db 17 VRIGYLPITDPTPLVAHNNGLFEABGIRKAPVLTLSMAQVIEAFISGVNTHLLSPM 76
 QY 101 PLAAVSGFTKADIEVPSMGFNCAITVSNELIWHQMKPNIPLEGKPVHPRIKADYLKPV 160
 Db 77 TWBAR-YGSKVPKVVAMNHVGGSLTVAPDI-----SAVKQLGKTV----- 118
 QY 161 VEKYKAEKGFNNAMTFPAGSHNIKIRYMLAAGINPGYISPPQDISGQIGADAL-LSVT 219
 Db 119 -----AIPWYSIHNVVLQQLNDGLTF--VSKP--ANAQLAANEVNLVL 161
 QY 220 PFPMPSTLEAGTIFGYCVGEPNMQAVFKIGVPIVITDEELMKD-----TPKXFGV 272
 Db 162 PPSDMPFALASKRIAGYVAEPFALAENLKVGRVQRFQDVMRNHACVVFMEHHLINN 221
 QY 273 TKQMAEKYPTVYLAVTKALIRAAIMLDADNNKRRKAEIMLAQ--KQYGVADYEVLAAS 329
 Db 222 RPEWSQK-----VNAIVKAQOW---TRDHTTEAALLSRGPKYTPHEPAVALTKV 270
 QY 330 M-----NGTFEYKDDKRALPDNTEFFRHGASYSYSAVYVLTQLRRMGMIN 377
 Db 271 LAPAEDRAGYIYSGAIRHQWDEKRI-DROPY-----PPSYTEE---LVKRLKTLTI- 320
 QY 378 EFKRDNMVYLDTAKNVYRPDIYLAARELV 406
 Db 321 --EGDNTFSLGLDPAY-----AARDLV 340

RESULT 14

ABU22033 ID ABU22033 standard; protein; 391 AA.

XX AC ABU22033;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #7560.

XX KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Burkholderia fungorum.

XX PN MO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KU, Zyskind JW,
PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX MPI; 2003-029926/02.
DR N-PSDB; ACA25903.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 49957; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIFO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 391 AA;

Query Match 9.0%; Score 220.5; DB 6; Length 391;
Best Local Similarity 24.4%; Pred. No. 2.6e-11;
Matches 94; Conservative 63; Mismatches 136; Indels 93; Gaps 16;

QY 9 SKKLLITTSASLAWGLTAPDVAGVAKLEKED---LKEFGIKLTDMAPLVAAKEGFF 64
DB 2 SREWMKASMFVAG--AGPLLGALNABAATDPNAPVIGYLPIDAPPLVAHNGGF 59
QY 65 EDEGFLVQLEA-QAMKVMMDRVVNGELDSHMLAPAPLAASVGFCTKADIEVPSMGFN 123
DB 60 EAAAGLAVERPTLLRSWAOLVEAFVLSQVWVHLLAPMTIWAR--YGSQAPAKVAVANNHN 117
QY 124 GNAITVSNIEIHWOMKNIPLEGSKPVHPKADYLKVEKYEKGFPMAMTPPGSHN 183
DB 118 GSALTIVSPIT---GKVADLGGKTV-----AVPEWSTHN 148
QY 184 IKLRWYLAAGINP-----GYSSPPODISGOIGADALLSVTPPPQMPSTLEAGTIPGYCV 238
DB 149 VVVQDMLRTQGLVPLVKKGDLKP-----NEVRLTVMAPSDMPALASRQINGFIV 199

QY 239 GEFPMNOAFKGIQGVITDEELMKD-----TPRKVGVYKQNAEKYPNTYLAATKL 291
DB 200 AEPFNAAAGLXIKGLFTGTGVWKNHACCVFMHERDLTERAAMQK-----VVDV 252
QY 292 IPRAIWLADNNKRNKEAIEMLA---QKQY-----VGADVEVLAASWNGTFEYKDDKRL 344
DB 253 VKRQLMTRA---HQPEAAGLSKQGNHTYHTAVALTEVLAAPAGDGRYLD--RAI 306
QY 345 -----PDFNTEFRHGASYPSS 361
DB 307 IHADWAKRIDFQPY-----PYPAYT 327

RESULT 15
AAB76701
ID AAB76701 standard; protein; 275 AA.
XX
AC AAB76701;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:384.
XX
XX Corynebacterium glutamicum; *brevibacterium lactofermentum*; MCT;
XX membrane construction and membrane transport protein; petroleum spill;
XX hydrocarbon degradation; gram positive aerobic bacterium; marker;
XX identification; microorganism; fine chemical production; transformation;
XX genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000MO-IB000926.
XX
PR 25-JUN-1999; 99US-0141031P.
PR 08-JUL-1999; 99DE-01031454.
PR 08-JUL-1999; 99DE-01031478.
PR 08-JUL-1999; 99DE-01031563.
PR 09-JUL-1999; 99DE-01032122.
PR 09-JUL-1999; 99DE-01032124.
PR 09-JUL-1999; 99DE-01032125.
PR 09-JUL-1999; 99DE-01032128.
PR 09-JUL-1999; 99DE-01032180.
PR 09-JUL-1999; 99DE-01032182.
PR 09-JUL-1999; 99DE-01032190.
PR 09-JUL-1999; 99DE-01032191.
PR 09-JUL-1999; 99DE-01032209.
PR 09-JUL-1999; 99DE-01032212.
PR 09-JUL-1999; 99DE-01032227.
PR 09-JUL-1999; 99DE-01032228.
PR 09-JUL-1999; 99DE-01032229.
PR 09-JUL-1999; 99DE-01032230.
PR 14-JUL-1999; 99DE-01032927.
PR 14-JUL-1999; 99DE-01033005.
PR 14-JUL-1999; 99DE-01033006.
PR 14-JUL-1999; 99DE-01040764.
PR 27-AUG-1999; 99DE-01040765.
PR 27-AUG-1999; 99DE-01040766.
PR 27-AUG-1999; 99DE-01040830.
PR 27-AUG-1999; 99DE-01040831.
PR 27-AUG-1999; 99DE-01040832.
PR 27-AUG-1999; 99DE-01040833.
PR 31-AUG-1999; 99DE-01041378.
PR 31-AUG-1999; 99DE-01041379.
PR 31-AUG-1999; 99DE-01041395.
PR 03-SEP-1999; 99DE-01042077.
PR 03-SEP-1999; 99DE-01042078.
PR 03-SEP-1999; 99DE-01042079.
PR 03-SEP-1999; 99DE-01042086.
XX

PA (BADI) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Pi XY

2000

DR N-PSDB; AAF67934.

PT Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying *C. glutamicum* or related bacteria, and as markers for
PT transformation.

PS Claim 20; Page 720-721; 1119pp; English.

[illegible]

CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention

SQ Sequence 275 AA;

Query Match	9.08;	Score 219.5;	DB 4;	Length 275;
-------------	-------	--------------	-------	-------------

Best Local Similarity 23.1%; Pred. No. 1.9e-11;
Matches 66; Conservative 52; Mismatches 127; Indels 41; Gaps 5;

42 LKRGFIKLDMAPIAVALAEKGFEEDEGLFYQLAQNKKVMDRVVNGELDGSHTLAP 101
 26 LTIGYVPIAGSAPIAIADALGLFKHGCVNTLKKYSGMSDLWTAYAEQLDVAHMLSP 85

26 LTIQVPIAGSAPIAIADALGLFKHGVNTLKKYSGWSDLWTAYATEQLDVAHMLSPMT 85

102 LAASVGF-GTKADIEVPFSMGFNGNAITVSNEIWHQMKPNI PLEGGKPVHPRIKADYLKPV 160

86 VAINAGVTNASRPTELTSFTONTNGQAITLASKHYGSVNSADIKG----- 130

161 VEKYKAEKPFENMAMTFPAGSHNIKLRWLAAGGINPGYSPQDISGQIGADALSTVP 220

31 ---MVLGIPFEYSV-----HALLLRDYLVSNAVDP-----IADLEIRLLR 167

QY 221 P P Q M P S T L E A G T I F G Y C V G E P W N Q Q A V F K G I G V P V I T D E L W K D T P E K V F G V T K Q W A E K Y 280

68 PADMAQLTVEGIDGFIGPFFNERAISNGSGRIWLTQKQWDKHPCCAVAMAKEWKA EH 227

QY 281 PNTYLAVTKALIRAIWLDADNNKRRKEAIEMLAQOYGVADVEVL 326

28 PTAQGVNLAEASAIL--SNPAQFSSARTLSQEKYLNQPATLL 271

Search completed: May 18, 2005, 17:42:07

Job time : 176 BECS

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2005, 17:32:50 ; Search time 27 Seconds
(without alignments)
1282.858 Million cell updates/sec

Title: US-10-689-200-2

Perfect score: 2443
Sequence: 1 MKTIRSSSKLLTLTSLA.....AKFAIGLKQTVAGKVD 464

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
4: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfill1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	577.5	23.6	456	US-09-489-039A-10363	Sequence 10363, A
2	446.5	18.3	474	US-09-540-236-2892	Sequence 2892, Ap
3	338.5	13.9	336	US-09-328-352-7572	Sequence 7572, Ap
4	292.5	12.0	249	US-09-252-991A-30501	Sequence 30501, A
5	240	9.8	419	US-09-252-991A-25707	Sequence 25707, A
6	219.5	9.0	275	US-09-602-787A-384	Sequence 384, App
7	157	6.4	197	US-09-252-991A-30392	Sequence 30392, A
8	121.5	5.0	336	US-09-902-540-12264	Sequence 12264, A
9	116.5	4.8	337	US-09-252-991A-26757	Sequence 26757, A
10	116.5	4.8	633	US-09-252-991A-25810	Sequence 25810, A
11	103.5	4.2	327	US-09-489-039A-12136	Sequence 12136, A
12	102	4.2	369	US-09-107-532A-5754	Sequence 5754, Ap
13	102	4.2	1062	US-09-397-550-4	Sequence 4, Appl1
14	102	4.2	1076	US-09-470-443-6	Sequence 6, Appl1
15	102	4.2	1082	US-09-397-550-5	Sequence 5, Appl1
16	102	4.2	1109	US-09-397-550-6	Sequence 6, Appl1
17	102	4.2	1115	US-09-397-550-23	Sequence 23, Appl1
18	102	4.2	1145	US-09-470-443-2	Sequence 2, Appl1
19	102	4.2	1145	US-09-470-443-4	Sequence 4, Appl1
20	102	4.2	1145	US-09-397-550-20	Sequence 20, Appl1
21	101.5	4.2	952	US-09-328-352-5611	Sequence 5611, Ap
22	101.5	4.2	1252	US-10-012-762-20	Sequence 20, Appl1
23	101.5	4.2	1252	US-09-704-036B-20	Sequence 20, Appl1
24	101	4.1	858	US-09-255-829-22	Sequence 22, Appl1
25	101	4.1	858	US-09-255-829-29	Sequence 29, Appl1
26	101	4.1	1169	US-09-255-829-20	Sequence 20, Appl1
27	98.5	4.0	315	US-09-248-796A-16091	Sequence 16091, A

28	98.5	4.0	1864	2	US-08-804-227C-3	Sequence 3, Appl1
29	98	4.0	324	4	US-08-956-171E-5209	Sequence 5209, Ap
30	98	4.0	324	4	US-08-781-986A-5209	Sequence 5209, Ap
31	96.5	4.0	448	4	US-09-198-452A-216	Sequence 216, App
32	96.5	4.0	457	4	US-09-438-185A-199	Sequence 199, App
33	96.5	4.0	659	1	US-08-258-639A-4	Sequence 4, Appl1
34	96.5	4.0	659	2	US-08-300-951A-4	Sequence 4, Appl1
35	96.5	4.0	659	5	PCT-US95-07391A-4	Sequence 4, Appl1
36	95.5	3.9	893	4	US-09-489-039A-14127	Sequence 14127, A
37	93.5	3.8	540	4	US-09-902-540-14003	Sequence 14003, A
38	93.5	3.8	1018	1	US-08-452-052-2	Sequence 2, Appl1
39	92	3.8	620	3	US-09-442-100-11	Sequence 11, Appl1
40	92	3.8	620	4	US-08-939-106-11	Sequence 11, Appl1
41	92	3.8	620	4	US-09-442-102-11	Sequence 11, Appl1
42	91	3.7	379	4	US-09-071-035-6	Sequence 6, Appl1
43	91	3.7	402	4	US-09-489-039A-11633	Sequence 11633, A
44	91	3.7	431	4	US-09-134-000C-6307	Sequence 6307, Ap
45	91	3.7	777	4	US-09-917-254-57	Sequence 57, Appl1

ALIGNMENTS

RESULT 1	US-09-489-039A-10363	Sequence 10363, Application US/09489039A
;	Patent No. 6610836	
;	GENERAL INFORMATION:	
;	APPLICANT: Gary Breton et al	
;	TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA	
;	FILE REFERENCE: 2709.2004001	
;	CURRENT APPLICATION NUMBER: US/09/489,039A	
;	PRIOR FILING DATE: 2000-01-27	
;	PRIOR APPLICATION NUMBER: US 60/117,747	
;	NUMBER OF SEQ ID NOS: 14342	
;	SEQ ID NO 10363	
;	LENGTH: 456	
;	TYPE: PRT	
;	ORGANISM: Klebsiella pneumoniae	
;	US-09-489-039A-10363	
Query Match	23.6%; Score 577.5; DB 4; Length 456;	
Best local Similarity	33.5%; Pred. No. 2.5e-49;	
Matches 153; Conservative 76; Mismatches 173; Indels 55; Gaps 17;		
QY	SSSKKLLTLTSLASLAVGLTIAPDV-----GAVGKLEKEDLKFGFIKLTDMAPLVAAB 60	
DB	44 SISRRRLTGAGALG--GAMTLPGVQAAMAGSDRPEQTRVAVGFPLTDCAPLAIAA 101	
QY	61 KGFEEBGLFVQLEAANMKVWMDRVNNGELDSHULAPAPLAASVFGTKADIEVP-FS 119	
DB	102 KGFDOXYGTTLVASKASMAVAVDKLVAGELDAHLTYGLYGLIEGIASKPOAMANLMT 161	
QY	120 MGENGAITVSNINIMQKPNIPLEGKPVHPKADYLKPVVEKYAKGKPFMMATFPA 179	
DB	162 IANNGAITSSTL--QEKVDTLGG-----LKLILR-SAPGS-YTFATHTFT 206	
QY	180 GSHNKLIRYLAAGGINSPPDIDISQIGADALLSVTPPOMBSTLEAGTIFGCVG 229	
DB	207 GTTAMLVYVLAAGIDP-----FNDVATVVVPPOMWNNRIGNNSGFCVG 253	
QY	240 EPNQAQVKGIGVPIITDELMKDPKPVFGYTKMAEKVPMTYLAVTALRAIIMLD 259	
DB	254 EPNNAALINDRIGFTATSDIPEPEKVLGTRRUVENNPPTARLVALLMEADRWI- 312	
QY	300 ADNNKRRKEAIEWLAKQYVGADEVYLAASMNSTFEYKODKALPDFN--TFRRHG-AS 356	
DB	313 AASPENTRETRARLARARCMINTKEQVLTGRMLG--EYDNGLGRRMKDAPHIRRWAGGEVS 370	
QY	357 YPSYSAVWYLTQLRWGMINEFKPDNWTYLDRAKVNYPDIYAAAKELVASGKAAEDF 416	

Db 371 FFWLSDGMWFLDQFRRMGLLKQ-APD--YLAVASRINRIDWQAAQ---AVGISA--- 421

QY 417 PADTSIKPSQNFIDKVPDPANPNYLAKEAIGLG 453

Db 422 PA---ARMRSSTLMGDTVMNSDPEGYARHFSIORKG 455

RESULT 2

US-09-540-236-2892

Sequence 2892, Application US/09540236

Patent No. 6673910

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATRA

TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2005-001

CURRENT APPLICATION NUMBER: US/09/540,236

CURRENT FILING DATE: 2000-04-04

NUMBER OF SEQ ID NOS: 3840

SEQ ID NO 2892

LENGTH: 474

TYPE: PRT

ORGANISM: M. catarrhalis

US-09-540-236-2892

Query Match 18.3%; Score 446.5; DB 4; Length 474;

Best Local Similarity 27.9%; Pred. No. 4.5e-36;

Matches 159; Conservative 82; Mismatches 180; Indels 71; Gaps 11;

QY 2 KTIIRSSSKLLTTSASIAVMGLTAPVGAAGVKL--EKEDLKFGFIKLTDMAPLAVA 59

Db 79 KTVGKTA-----LAIASVLPALQLQAAIDTLKPEKSDVIGFLPLCTPLIMAD 132

QY 60 EKGFEDEGLFVQLQAQANMKVMDRVNGLDGSIMLAPAPLASVGFGT-KADIEVP 118

Db 133 PLGYAEQGIKANILTRAGMALVRDQMRRELDAAFLAPMFAIMIGLSAKQNKVAA 192

QY 119 SMGFNGNATTVSNELIWHQKKNPRLGKGVPHIKADYLKPVVEKTKAGKPPNMAITFP 178

Db 193 TONTNQAALVMA-----LKHNNRNPKNWKGTFTALPFE 226

QY 179 AGSHNKLKRWLAAGINPGYSPQDISQIGADALLSVTPPQMPSTLEAGTIGYCV 238

Db 227 HSHHNYLAFYLAHGLD-----DKDKLALTPPRDMIAHLAKGNIDGFFG 273

QY 239 GEPNQAQVFKGIGVPIITDEELMKDTPKGVGTQMAKTPNTYLAATKALIRAIWL 298

Db 274 PEPFQRAVWDKAGYHTLSRDIMNGHPCCSFQTSQSFINDYQOTFLAMYRAIKANVM- 332

QY 299 DADNNKREALEMLAQKYGVADGVLAASNGTFEYKDKKALPDNTPFRHGA-S-Y 357

Db 333 -ANKDSIRKDLSPANYLNOPELVILQSIWGRPADGVDTIQVDP-----RMGFDM 386

QY 358 PSYSSAVWYLTQLRKGMINEFAPDNWYLDTKANVVRPDIYLAALAEKAKADEPP 417

Db 387 PWSVAAMWMTQMKKGYIT---GNINQDIANQV---MLDAKQOMQMGTYVQDEP 440

QY 418 ADTSIKPSQNFIDKVPDPANPNYLAKEAIGLKQTFVAG 459

Db 441 KITVMGKQ-----FNATDPDAYLDSFAIGHKSTGRLHG 473

RESULT 3

US-09-328-352-7572

Sequence 7572, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 7572

LENGTH: 336

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-7572

Query Match 13.9%; Score 338.5; DB 4; Length 336;

Best Local Similarity 26.5%; Pred. No. 2e-25;

Matches 96; Conservative 63; Mismatches 144; Indels 59; Gaps 9;

QY 34 VGKLEKEDLKFGFIKLTDMAPLAVAEEKGFDEGLFVQLBAQANMKVMDRVNGLDQ 93

Db 1 MSKLEKTOQLGFIPLDICALIMAKQGFEEVDLTVLKEASASLRDLAPLDA 60

QY 94 SHMLAPAPLASVGFCTKAD-----IEVPSNGFNATVSNELIWHQ---KPNIPPE 144

Db 61 AHCLSMLPAAAMG-----ADQIGIALQDTPLVLSKNRAFLISQKLIHQALIKENDVAQTT 116

QY 145 GSKPVHPIKADYLKPVVEKTKAGKPPNMAITFPAGSHIKLRMYLAAGINPGYSPQ 204

Db 117 AOKVLIQYIEODH-----TSLAHVFKSHIHHYCLREWTALA----- 152

QY 205 DISQIGADALLSVTPPQMPSTLEAGTIGYCVGEPNQAQVFKGIGVPIITDELMK 264

Db 153 --DSRLAQTLKALPFPYMWALDNHVIDGFCVGEPMTOGELGLSKIVCSQDIIFN 210

QY 265 TPEKYGVGTQMAEKYPTTYLAATKALIRAIWLDADNNKREALEMLAQ---KQYVG 320

Db 211 VADKVLAVTQEWAEQHPTLVALTALIMKAQ--KELSNLKDPAPVYKLLVEGIVRFHS 268

QY 321 ADVEYLAASNGTFEYKDKKALPDNTPFRHGA-SYPSYSAWYLTQLRKGMINEK 380

Db 269 EEVHNDKRYMIQNIYKHLVKNAAQPPEDFH-----MLFOQMORKEKL-QLA 314

QY 381 PD 382

Db 315 PD 316

RESULT 4

US-09-252-991A-30501

Sequence 30501, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 30501

LENGTH: 249

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30501

Query Match 12.0%; Score 292.5; DB 4; Length 249;

Best Local Similarity 31.3%; Pred. No. 5.4e-21;

Matches 73; Conservative 40; Mismatches 105; Indels 15; Gaps 6;

QY 218 VTPPQMPSTLEAGTIGYCVGEPNQAQVFKGIGVPIITDEELMKDPEKFGVTKQMA 277

Db 19 VVPQQWGHLOAGRIDGCGAPWGAALVDDGQGFATATSGAIPWDHEKVLGTTAFV 78

QY 278 EKPNTYLAATKALIRAIWLDADNNKREALEMLAQKYGVADGVLAASNGTFEY 337

Db 79 DAYPNTALALVMAVLDASRFIE-QVAENRLLGTALISGRDYDAPALQAPFAFFGKYQDG 137

QY 338 KODKRALPDENTFRHG-ASYPSYSSAVYLTOLRRKGMINEKPDNWTDTAKNTYRPD 396
DB 138 LGNAWODPPLRFYAGEVVRPWLSDGMFMETQFRRWGLRRE-DPD--YIGIARVQOTA 194
QY 397 IYLAAXEVAEKADEPDAOTSIKPSQNFPLDKVFPDANKPNDLAKFAI 449
DB 195 LYRDAATLGL--RLDGADMRST-----LIDGRTWGDSPAGYARSFPI 237

RESULT 5
US-09-252-991A-25707
; Sequence 25707, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25707
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25707

Query Match 9.8%; Score 240; DB 4; Length 419;
Best Local Similarity 24.6%; Pred. No. 2,5e-15;

Matches 93; Conservative 68; Mismatches 141; Indels 76; Gaps 14;

QY 9 SKKLLITLSASLAWG-LTAPDVGAVKLEKE-DLKEFGTKLTMAPLAVALAEKGFED 66
DB 29 SRRDIKTLAALASAGALPLSLSLQARAAEPDAPVIGYLPITDAPPLIVAHANGIFEA 88
QY 67 EGIFFVLEQA---NMKVWMDRVNGELDGSMLAPPLASVGFQTKDIEVPFSGFN 123
DB 89 EG-IOAEPVTLRSMAQVLEAFISQOVNTHILSPVTWAR--YSSKVPKAVVANNHG 144
QY 124 GNAITVSNIEWMQKNIPLGGKPVPIKADYLKPEVYKAEKGFPMNAATFPAGSHN 183
DB 145 GSGTLVAPEL-----ADVQLGGKSV-----AIFWYSIHN 175
QY 184 IKLRVYLAAGNGINPGYSPDQISQIGA-----DALLSVTPPPQWSTLEAGTIFGYCV 238
DB 176 VVLQQLLRNGL-----RAVSRAVGAALADEVNLVLPSPDMPALASKRHGYIV 227
QY 239 GEPWNOAVFKGIGVYITDEELMKTPEKVFQVTKQMAEKYNTYLAATKALIRAAIYL 298
DB 228 ASEPFLALNLTQVQVQFTGDVWRNHACCVFMHEDERRPQWSQKVNAIVKQLW- 286
QY 299 DADNNGRKEAIEMLAQ--KQYVADVEYL-----AASNGTFEYKODKRA 343
DB 287 ---TTRHRAVAQAQLLKAGANRTPPAPEVLGAVLAPGAEQQAAYLASGIRHADWOER 343
QY 344 LPPDFTFFRHGASYPYS 361
DB 344 I-DFQPY-----PYPSYT 355

RESULT 6
US-09-602-787A-384

; Sequence 384, Application US/09602787A

; Patent No. 6696561

; GENERAL INFORMATION:

; APPLICANT: Pompeius, Mark

; APPLICANT: Krüger, Burkhard

; APPLICANT: Schöder, Hartwig

APPLICANT: Zelder, Oskar
APPLICANT: Habethauer, Gregor
TITLE OF INVENTION: CORNEAECTERIUM GLUTAMICUM GENES ENCODING PROTEINS
TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
FILE REFERENCE: BGI-125CP
CURRENT APPLICATION NUMBER: US/09/602,787A
CURRENT FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: USN 60/141031
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: DE 19931454.3
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931478.0
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931563.9
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19932122.1
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932124.8
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932125.6
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932128.0
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932180.9
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932182.5
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932190.6
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932191.4
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932209.0
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932212.0
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932227.9
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932228.7
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932229.5
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932230.9
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932927.3
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19933005.0
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19933006.9
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19940764.9
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940765.7
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940766.5
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940830.0
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940831.9
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940832.7
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940833.5
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19941378.9
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: DE 19941379.7
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: DE 19941395.9
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: DE 19942077.7
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: DE 19942078.5

Best Local Similarity 22.5%; Pred. No. 0.0051;
Matches 79; Conservative 52; Mismatches 157; Indels 63; Gaps 13;

QY 13 LTLTSLAAGVGLTIPDVAGVGLKEKEDLKFGFIKLTDMAPLAIAEKGFEEGLFVQ 72
DB 8 LRLSLNTLLAAGLAVTLFSPILASQAEEKIKGVVAVAGVGFVADKDLFPHGIDVD 67
QY 73 LEAANWKKVMDRVNVELDGHMLAPLAASVGFCTKADIEVPSMGNGNAITVSN 132
DB 68 LQFNDPALPTAMVSSALDGG-MLTYDQVAVAKGLKRRVMPIDFSNGDAIV- 122
QY 133 IMHGMKNIPLEGKPHPIKADYLKVEKYKAEGPFPMAMTFPPGSHNITKRTYLA 132
DB 123 -----ADASIQSIADFKGVGFN-----PLSPSDFLLAAYALQ 156
QY 193 GGINPGYSPPODISGQIGADALLSVTPPOMPSTLEAGTIFGCVSEPPNQOAVFGIG 252
DB 157 NGMS-----DKDMA-----VNMT-PEGIGPAMASGMLPGVTYEPVNSQILSMGG 202
QY 253 --VPVITDELMKDTPEKVFQVTKQMAKKYPTTYLAATKALIRAAI--WLD--ADNNKNR 306
DB 203 DKFKVYSS--KDAFGLITDYL-----VPEDAVIAKKPAIKAMIGYLDGLAYMQAHP 254
QY 307 KEALIEMLAKQVYGAD--VEVLAASMGTEYEKEDKRALPDRTFFRGA 355
DB 255 EESAKITIGDVLGVSAAEAVQOMAGAVN--IPLAEMGKSTFGDDTHSFHS 303

RESULT 10
US-09-252-991A-25810
Sequence 25810, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 25810
LENGTH: 623
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25810

Query Match 4.8%; Score 116.5; DB 4; Length 623;
Best Local Similarity 21.3%; Pred. No. 0.014;
Matches 93; Conservative 55; Mismatches 169; Indels 119; Gaps 18;

QY 60 EKGFFED---EGLFVQLEAQAAMKVMKRVNVELDGHMLAPL-----AASVFGT 110
DB 201 ELGLLEKYLPRDGRYXADYRVEK-----NFTSGAPLTNEMVAGKLDFGA 246
QY 111 KADIEVPSMGNGNAITVSN-----ITVSNETWHQMKPIPLEGKPHPIKADYIKPV 160
DB 247 MAD---FQTFNGVAPAAAGKSLFISVLSGSTRSGNGIIVPSSGVPVOST----- 294
QY 161 VEKYKAEGKPFPMAMTFPPGSHNITKRYMLAAGINGVYSPPODISGQIGADALLSVTP 220
DB 295 -----AEKGGTISVFPASAHGMLRAVAAG-----WDPRDVT-----IIAP 335
QY 221 PPOMPSTLEAGTIFGCVSEPPNQOAVFGIGVPIITDELMKDTPEKVFQVTKQMAKKY 280
DB 336 PEIAGSALQAMKIAAHADFVPFAELFPNRRGFARKIYVSGQANAPTFHGAL-VDAAYERY 394
QY 281 PNTYLAATKALIRAAIWLADNNKRRKALIEMLAKQVYGADEV----- 325
DB 395 PEVAVAYLRASIEANRLAEPK-YSELIE-----KVTGIEABVNYLFHGPLGLQTRDQ 448

QY 326 -----LAASMGTFEYKDKKALPDENTF-----FRHASYSYS 362
DB 449 AMREYQALATSID-TLKLLKKADRG-L-DVERFVDRYIREAVRQAGLDYQALADYAP 506
QY 363 AWTYTLQRRKWNINEFK--PDWNYLDTA--KRVYRPDIYLAAKEVYASGA-----KAE 414
DB 507 SPLRASPARSGRPIIEEFERVAQIIVWGEARVRYASPSALADLAELKAGKALIRAYQ 566
QY 415 DFPADTSIKPSQNF 430
DB 567 DRDSGIKLLANQAMFV 582

RESULT 11
US-09-489-039A-12136
Sequence 12136, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 12136
LENGTH: 327
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12136

Query Match 4.2%; Score 103.5; DB 4; Length 327;
Best Local Similarity 20.3%; Pred. No. 0.099;
Matches 65; Conservative 53; Mismatches 137; Indels 65; Gaps 12;

QY 19 SLAWGTLTIPDVAGVGLKEKEDLKFGFIKLTDMAPLAIAEKGFEEGLFVQLEAQA 78
DB 14 ALTLGHLFSSIGHAAPADALRIGYQKGSMTLA-----KSHQLEQRYQTH 65
QY 79 WKVMDRVNVELDGHMLAPLAASVGFCTKADIEVPSMGNGNAITVSNIEHQM 138
DB 66 ISWI-----EPPAGQMLBALVNSIDIGTDPPIPAQAADLVYGAE--PRK 115
QY 139 PN---IPLEGKPHPIKADYLKPVVEKYKAEGKPFPMAMTFPPGSHNITKRTYLAAG-- 193
DB 116 PKAEVILVAGSPRIHV-----TELKGGKVAFOKSSSHNITLRLAQLAGLK 162
QY 194 --GINPGYSPPODISGQIGADALLSVTPPOMPSTLEAGTIFGCVSEPPNQOAVFGI 251
DB 163 FSDIQEYTLAP-----ADA-----RAAFQGSNVAMALWDPYSSALLQG- 202
QY 252 GVEVITDELMKDTPEKVFQVTKQMAKKYPTTYLAATKALIRAAIWLADNNKRRKAE 311
DB 203 GARVLDTGDTLQGT-GSFYLASRPVABRG-----AFIEGLVDTFTQADALTHSQASIT 257
QY 312 MIAQKQVYGADEVVLAASMN 331
DB 258 LIAKT--MGHPAIVIASYLD 275

RESULT 12
US-09-107-532A-5754
Sequence 5754, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:

```

ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051,571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5754:
SEQUENCE CHARACTERISTICS:
LENGTH: 369 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) LOCATION 1..369
SEQUENCE DESCRIPTION: SEQ ID NO: 5754:
US-09-107-532A-5754

Query Match
Best Local Similarity 21.5%; Pred. No. 0.17; Length 369;
Matches 85; Conservative 47; Mismatches 138; Indels 126; Gaps 20;

QY 10 KLLLTLSASLAWGTLTAPDVGA---VKELEKEDIKFGFIKLTMDAPLAAVAEKGFED 66
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 41 KKLWLLPILLISAGTAKETSSKOEIKDKKVTLLVDYPTNHTGIYLAKEKGYKE 100
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 67 EELFVQL-----EAQAMKVVMDRVVNGELDSHMLAPPLAASV---- 106
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 101 AGLVNOIIEPGENSTISGLVGADKAQFGVSYOED--VTVAHAGQNT--PVKAIATATIKIN 157
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 107 --GFGTKADIEVPFSGFNAGNAITVSEIMHQMKPNIPLEGKPVPIKADYIKPVYKY 164
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 158 TSGFATLSNSNHSHPDFEGKTYAG---WQS-----PSEEAIVAKAMEK- 198
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 165 KAEGRFNNAMTFPAGSHNIKLRVYLAAGINPGYSPPODISGQI--GADALLSVTPPO 223
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 199 --DGGDFS-----KLT--WVGSNGEGBESLGKSDIQWYREBMD----- 233
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 224 MSTLEAGTIFGVCVGEPMNQAVFEGIGVPI--ITDEELMKOTPEKVFQVTKOMAEKYN 282
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 234 MKAKEAGLEIHWYIPLKEIDERLDYV---TPVITMDQIKSDPELVY-----Q 278
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 263 TLAVATKALIRAIWMDADNNKORKEAIEMLAQOVVADVEVLASNMNGTFEYKODKR 342
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 279 SEMDITKGYQAL--KQPNDSAK-----LQK-----YAKENDR 311
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 343 ALPDENTFRHAGASYPSSAVWYLTOLRRQWINE 378
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 312 -----TLEESQAFISKV-----YTDDPENKGLME 337
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 13
US-09-397-550-4
; Sequence 4, Application US/09397550
; Patent No. 6783952
GENERAL INFORMATION:
APPLICANT: Warner-Lambert
TITLE OF INVENTION: Secreted soluble alpha2 delta-2, alpha2 delta-3, alpha2
TITLE OF INVENTION: delta-4 calcium channel subunit polypeptides and
TITLE OF INVENTION: screening assays using same
FILE REFERENCE: 180
CURRENT APPLICATION NUMBER: US/09/397,550
CURRENT FILING DATE: 1999-09-16
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 1062
TYPE: PRT
ORGANISM: Homo sapiens
US-09-397-550-4

```

```

Query Match
Best Local Similarity 23.7%; Pred. No. 0.95; Length 1062;
Matches 79; Conservative 42; Mismatches 142; Indels 70; Gaps 17;

```

```

QY 7 SSKKLLTSLASLAWGTLTAPDVGA---VKELEKEDIKFGFIKLTMDAPLAAVAEKGFED 66
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 287 SSPKQNVITVDVSGSVGLTL-----KLMKTSVCEMLDITLSDDYNNVASFNEKAP 338
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 67 EELFVQL--EAQAMKVVMDRVVNGELDSHMLAPPLAASVGFRTKADIEVPFSGMFGCN 125
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 339 VSCFTHLVQAVNRKVKFKEAVQG-----WAKGTGYKAGFEYVFDQLQNSN 386
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 126 AITVSEIMHQMKPNIP--EGKPVPIKADYIKPVYKAEKGFNNAMTFPAGSHN 183
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 387 -ITRAN-----CNKMIIMPTDGB-----DRVDVEFKYMWPIRYTVV--PTFSVGQHN 432
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 184 IKLR--YMLAAGINPGYSPPODISG--QIGDALLSVTPPOMSTLEAGTIFGVCVE 240
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 433 YDVTPLQWMAA--NKGYFEIPISGAIRINTQEVYLDVGRPMVLAKGKAKV----- 483
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 241 PMNQAVFK--GIGVPIITDEELMKOTPEKVFQVTKOMAEKYNPTVLAATKALIRAIWMD 299
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 484 QMTN--VIEDALGLGLVVTGTL-----PVNLTQDQGEPEKKN-----QLITGMGID 528
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 300 ADNNKRRKEAIEMLAQOVVADVEVLASNMNG 332
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 529 VALND-----IKRLTPNVTILGANGVVFALDING 556
DB |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 14
US-09-470-443-6
; Sequence 6, Application US/09470443
; Patent No. 6441156
GENERAL INFORMATION:
APPLICANT: Ierman, Michael I.
APPLICANT: Minna, John D.
APPLICANT: Latif, Farida
APPLICANT: Wei, Ming-Hui
APPLICANT: Sekido, Yoshitaka
APPLICANT: Gao, Boning
APPLICANT: Duh, Ruh-Mei
TITLE OF INVENTION: Calcium Channel Compositions and Methods of Use Thereof
FILE REFERENCE: NIH-05043
CURRENT APPLICATION NUMBER: US/09/470,443
CURRENT FILING DATE: 1999-12-22
EARLIER APPLICATION NUMBER: 60/114,359
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 6
LENGTH: 1076
TYPE: PRT

```

ORGANISM: Homo sapiens
US-09-470-443-6

Query Match 4.2%; Score 102; DB 4; Length 1076;
Best Local Similarity 23.7%; Pred. No. 0.97;
Matches 79; Conservative 42; Mismatches 142; Indels 70; Gaps 17;

QY 7 SSSKKLLLTLSASLAWGLTIAPDVAGVKLEKEDLKFGFIKLTDMAPLAVAAEKGFED 66
DB 218 SSPKDWIIVDVSGSVGLTL-----KLMKTSVCEMLDTISDDYVNVASFNEKAQP 269
QY 67 EGI.FVOL- EAQANWKYVMDRVVNGELDSIMLAPPLAASVGFRTADIEVPPSMGFNGN 125
DB 270 VSCFTHLVQANVANKKVFKEAVOG-----MVAKGTTGYKAGFEYAFDQLNSN 317
QY 126 AITVSEIHWOMKPNIP.L- EGSKPVHPIKADYIKPVEKYKABGKPFNNAMTFPAGSHN 183
DB 318 -ITRAN-----CNKMTIMPTDGE-----DRVDVFETKNNPNRTVRV-FTFSVGQHN 363
QY 184 IKLR--YMLAAGINPGYSPPODISG-OIGADALISVTPPPQMPSTLEAGTIFGYCVGE 240
DB 364 YDVTPLQWACA--NKGYPFIRPSIGAIRINTQEVLDVGRPMVLAGKEAKQV----- 414
QY 241 PMNQAVFK-GIGVPVITDELMKDTPEKVFQTKQAEKYPNTYLAVTKALIRAAIWL 299
DB 415 QWTN--VYEDALGLVVTGTL-----PVFNLTODGPEKKN-----QLILGWGID 459
QY 300 ADNKKRKEAIEMLAQKQYVGADEVYLAASMG 332
DB 460 VALND-----IKRLTPNTLIGANGYVFAIDNG 487

RESULT 15
US-09-397-550-5
Sequence 5, Application US/09397550
Patent No. 6783952
GENERAL INFORMATION:
APPLICANT: Warner-Lambert
TITLE OF INVENTION: Secreted soluble alpha2 delta-2, alpha2 delta-3, alpha2
TITLE OF INVENTION: delta-4 calcium channel subunit polypeptides and
FILE REFERENCE: 180
CURRENT APPLICATION NUMBER: US/09/397,550
CURRENT FILING DATE: 1999-09-16
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 5
LENGTH: 1082
TYPE: PRT
ORGANISM: Homo sapiens
US-09-397-550-5

Query Match 4.2%; Score 102; DB 4; Length 1082;
Best Local Similarity 23.7%; Pred. No. 0.98;
Matches 79; Conservative 42; Mismatches 142; Indels 70; Gaps 17;

QY 7 SSSKKLLLTLSASLAWGLTIAPDVAGVKLEKEDLKFGFIKLTDMAPLAVAAEKGFED 66
DB 287 SSPKDWIIVDVSGSVGLTL-----KLMKTSVCEMLDTISDDYVNVASFNEKAQP 338
QY 67 EGI.FVOL- EAQANWKYVMDRVVNGELDSIMLAPPLAASVGFRTADIEVPPSMGFNGN 125
DB 339 VSCFTHLVQANVANKKVFKEAVOG-----MVAKGTTGYKAGFEYAFDQLNSN 386
QY 126 AITVSEIHWOMKPNIP.L- EGSKPVHPIKADYIKPVEKYKABGKPFNNAMTFPAGSHN 183
DB 387 -ITRAN-----CNKMTIMPTDGE-----DRVDVFETKNNPNRTVRV-FTFSVGQHN 432
QY 184 IKLR--YMLAAGINPGYSPPODISG-OIGADALISVTPPPQMPSTLEAGTIFGYCVGE 240
DB 433 YDVTPLQWACA--NKGYPFIRPSIGAIRINTQEVLDVGRPMVLAGKEAKQV----- 483
QY 241 PMNQAVFK-GIGVPVITDELMKDTPEKVFQTKQAEKYPNTYLAVTKALIRAAIWL 299

DB 484 QWTN--VYEDALGLVVTGTL-----PVFNLTODGPEKKN-----QLILGWGID 528
QY 300 ADNKKRKEAIEMLAQKQYVGADEVYLAASMG 332
DB 529 VALND-----IKRLTPNTLIGANGYVFAIDNG 556
Search completed: May 18, 2005, 17:42:39
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2005, 17:39:18 ; Search time 144 Seconds
(without alignments)
1077.858 Million cell updates/sec

Title: US-10-689-200-2
Perfect score: 2443
Sequence: 1 MKTIIRSSSKLLTLTSLASL.....AKFAIGLKQGVAGKVVDD 464

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1434725 seqs, 334507595 residues
Total number of hits satisfying chosen parameters: 1434725

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: Published Applications_AA.*
 - 2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2443	100.0	464	16	US-10-689-200-2
2	525	21.5	403	15	US-10-282-122A-69975
3	240	9.8	399	15	US-10-282-122A-66370
4	240	9.8	399	15	US-10-389-647-502
5	234.5	9.6	396	15	US-10-282-122A-49807
6	227	9.3	396	15	US-10-282-122A-44527
7	226	9.3	394	9	US-09-738-626-4906
8	222.5	9.1	372	15	US-10-282-122A-68363
9	220.5	9.0	391	15	US-10-282-122A-49957
10	219.5	9.0	375	15	US-10-627-476-384
11	217.5	8.9	391	15	US-10-282-122A-48182
12	144	5.9	342	15	US-10-282-122A-72603
13	140	5.7	192	15	US-10-282-122A-47676

14	116.5	4.8	883	14	US-10-156-761-8511	Sequence 8511, App
15	111.5	4.6	333	15	US-10-156-761-9969	Sequence 9969, App
16	108	4.4	533	15	US-10-425-114-51223	Sequence 51223, A
17	107	4.4	518	15	US-10-424-599-198653	Sequence 198653, A
18	106.5	4.4	461	15	US-10-282-122A-46659	Sequence 46659, A
19	105	4.3	419	15	US-10-282-122A-46437	Sequence 46437, A
20	105	4.3	1130	15	US-10-275-595A-6	Sequence 6, Appl
21	105	4.3	3352	14	US-10-156-761-7961	Sequence 7961, App
22	105	4.3	4455	15	US-10-287-226-304	Sequence 304, App
23	104	4.3	866	14	US-10-241-596-104	Sequence 104, App
24	104	4.3	1420	14	US-10-241-596-110	Sequence 110, App
25	103.5	4.2	280	9	US-09-815-242-11714	Sequence 11714, A
26	103	4.2	864	14	US-10-241-596-102	Sequence 102, App
27	102.5	4.2	320	15	US-10-369-493-18239	Sequence 55774, A
28	102	4.2	613	15	US-10-282-122A-55774	Sequence 18239, A
29	102	4.2	860	14	US-10-241-596-175	Sequence 175, App
30	102	4.2	862	14	US-10-241-596-94	Sequence 94, App
31	102	4.2	866	14	US-10-241-596-88	Sequence 88, App
32	102	4.2	867	14	US-10-241-596-96	Sequence 96, App
33	102	4.2	867	14	US-10-241-596-98	Sequence 98, App
34	102	4.2	870	14	US-10-241-596-92	Sequence 92, App
35	102	4.2	871	14	US-10-241-596-84	Sequence 84, App
36	102	4.2	871	14	US-10-241-596-86	Sequence 86, App
37	102	4.2	871	14	US-10-241-596-90	Sequence 90, App
38	102	4.2	1062	17	US-10-902-531-4	Sequence 4, Appl
39	102	4.2	1076	14	US-10-116-949-6	Sequence 6, Appl
40	102	4.2	1082	17	US-10-902-531-5	Sequence 5, Appl
41	102	4.2	1109	17	US-10-902-531-6	Sequence 6, Appl
42	102	4.2	1115	17	US-10-902-531-23	Sequence 23, Appl
43	102	4.2	1145	14	US-10-116-949-2	Sequence 2, Appl
44	102	4.2	1145	14	US-10-116-949-4	Sequence 4, Appl
45	102	4.2	1145	17	US-10-482-029-114	Sequence 114, App

ALIGNMENTS

RESULT 1
US-10-689-200-2
; Sequence 2, Application US/10689200
; Publication NO. US20040126848A1
; GENERAL INFORMATION:
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Ni, Hao
; APPLICANT: Ye, Rick
; APPLICANT: Picataggio, Stephen
; APPLICANT: Wang, Tao
; APPLICANT: Seip, John E.
; TITLE OF INVENTION: NATURAL PROMOTERS FOR GENE EXPRESSION IN CL METABOLIZING BACTERIA
; FILE REFERENCE: CL1747 US NA
; CURRENT APPLICATION NUMBER: US/10/689,200
; PRIOR FILING DATE: 2003-10-20
; PRIOR FILING DATE: 2002-10-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Methylobacillus sp. 16a
; US-10-689-200-2

Query Match 100.0%; Score 2443; DB 16; Length 464;
Best Local Similarity 100.0%; Pred. No. 2.5e-216;
Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTIIRSSSKLLTLTSLASLAWGLTIADVGAAGKLEKEDLKFRTKLTDAAPLAAAE 60
1 MKTIIRSSSKLLTLTSLASLAWGLTIADVGAAGKLEKEDLKFRTKLTDAAPLAAAE 60
Db 1 MKTIIRSSSKLLTLTSLASLAWGLTIADVGAAGKLEKEDLKFRTKLTDAAPLAAAE 60
QY 61 KGFPEDEGLFVQLEADQANKVMVDRVVGELDGSMLAPAPLAASVGFRTKADIEVPPSM 120
61 KGFPEDEGLFVQLEADQANKVMVDRVVGELDGSMLAPAPLAASVGFRTKADIEVPPSM 120
Db 61 KGFPEDEGLFVQLEADQANKVMVDRVVGELDGSMLAPAPLAASVGFRTKADIEVPPSM 120


```

QY 121 GENGNAITVSNELTHQMKPNIPLEGKPVHPRIKADYLKPVVEKXKAEKGFPMMAATPPAG 180
DB 121 GENGNAITVSNELTHQMKPNIPLEGKPVHPRIKADYLKPVVEKXKAEKGFPMMAATPPAG 180
QY 181 SHNIKRLVYLAAGGIPGYSPPODISGOIGADALSTYTPPPMPSTLEAGTIFGCVSE 240
DB 181 SHNIKRLVYLAAGGIPGYSPPODISGOIGADALSTYTPPPMPSTLEAGTIFGCVSE 240
QY 241 PNNQAVFKGIVGVITDEELMKDTEKXVFGVTKQMAEKYPTNTYLAVTKALIRAAIWLDA 300
DB 241 PNNQAVFKGIVGVITDEELMKDTEKXVFGVTKQMAEKYPTNTYLAVTKALIRAAIWLDA 300
QY 301 DNNKPKKALIEMLAQOYVIGADVYLAASMGFFEEKDKKALDPNTFFRGASVPSY 360
DB 301 DNNKPKKALIEMLAQOYVIGADVYLAASMGFFEEKDKKALDPNTFFRGASVPSY 360
QY 361 SSAVWYLTQRRGMINEFKPDNMYLDTAKNVPRIYLAAKEVLAEGKAKAEDPADT 420
DB 361 SSAVWYLTQRRGMINEFKPDNMYLDTAKNVPRIYLAAKEVLAEGKAKAEDPADT 420
QY 421 SIKPSQNFIDKVPDPANKPNDYLAKEFALGLKGQTVAGSKVVD 464
DB 421 SIKPSQNFIDKVPDPANKPNDYLAKEFALGLKGQTVAGSKVVD 464

```

RESULT 2

```

US-10-282-122A-69975
/ Sequence 69975, Application US/10282122A
/ Publication No. US20040029129A1

```

GENERAL INFORMATION:

```

/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Foreyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: Patent in version 3.1
/ SEQ ID NO 69975
/ LENGTH: 403
/ TYPE: PRT

```

```

/ ORGANISM: Pseudomonas syringae
/ US-10-282-122A-69975

```

```

Query Match 21.5%; Score 525; DB 15; Length 403;
Best Local Similarity 31.7%; Pred. No. 26-39;
Matches 132; Conservative 72; Mismatches 164; Indels 48; Gaps 12;

```

```

QY 38 EKEDLKFGFIKLTDMAPLVAANEKGFEEDEGLFVQLEADAMKVMYMDRVYNGELDGSNML 97
DB 19 ENMSLDVGFMAITDCAPLVVAATQGFAPQYGLSLMKRQTSWAGLDRIVSQLOQAHSI 78
QY 98 AAPLAASVGF--GTRKADIEVPFSGMGNALTVSNELTHQMKPNIPLEGKPVHPRIKAD 155
DB 79 YGLIYAVEIGIGSGPRTDAIILGLNONGCINTLSRE-----LDDAGVITPEALD 128
QY 156 YLKPVEKXKABKGFPMMAATPPAGSHNIKLYWYLAAGGIPGYSPPODISGOIGADAL 215
DB 129 -----KRAHSGSGSLTKFTQTEPFGNHAMWLYWYLAASQIHP-----LDDV 168
QY 216 LS-VTPPPMPSTLEAGTIFGCVSEPMNQAVFKGIVGVITDEELMKDTEKXVFGVTK 274
DB 169 TSVVVPPTQMAOHLQAGRIDGFCVSEPMGASAVQOQLGFTMATSOAIWPDHFGKVLGCTR 228
QY 275 QMAEKYPTNTYLAVTKALIRAAIWLDA DNNKPKKALIEMLAQOYVIGADVYLAASMGTF 334
DB 229 EYEQNPNTARALIMAVLEASRFIE-QSDHNRSTRQULSGVDYLDASLDCTEPRLLGQY 287
QY 335 EYKDKKRALDPNTFFFRHG-ASYPSSAVWYLTQRRGMINEFKPDNMYLDTAKNVP 393
DB 288 SDGLGNQWMDPRAVSFHDQGVNYPWMLSDGMFMFMQFRRWGLRE-DPD-VLAVASRVQ 344
QY 394 RPDYLAAKEVLAEGKAKAEDPADTSLKPSQNFIDKVPDPANKPNDYLAKEFAL 449
DB 345 QLDLYROANAL-----GIDAPSAT-LRSSQ--LIDGKVDGSDPDAGYARSFYK 390

```

RESULT 3

```

US-10-282-122A-66370
/ Sequence 66370, Application US/10282122A
/ Publication No. US20040029129A1

```

GENERAL INFORMATION:

```

/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Foreyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22

```



```

; SOFTWARE: Patentin version 3.1
; SEQ ID NO 49807
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
; US-10-282-122A-49807

Query Match
Best Local Similarity 23.3%; Score 234.5; DB 15; Length 396;
Matches 105; Conservative 77; Mismatches 150; Indels 119; Gaps 18;

QY 9 SKKLLTLLTSASLAWMLGTLTAPDVGAAGKLEKED---LKFGFKLTDMAPLAAAKGFF 64
DB 7 SREEMTKLASMTYAG--AAPLLAALNAAANRPDAPVIGYPTITDAAPLVAAHNGYF 64
QY 65 EDEGLFVQLA-QANRKYVMDRVVNGELDSHMLAPLAASVFGTKADIEVFGMGN 123
DB 65 DASGLAVEKPTLLRSWAQVLEAFISQVNVVHLLAPMTIAR--YSSQAPAKVAMNHYN 122
QY 124 GNAITVSNITWIMQMKNIPLBEGKRPVPIKADYLYKVEKYAKEGKPPKAMTFPAGSHN 183
DB 123 GSALTVAPI-----GKLELGGKTV-----AAPEWYSIHN 153
QY 184 IKLRWYLAAGINPGYSPPODISGOIGADAL-LSYTPPOMESTLEAGTIFGVCVGBPM 242
DB 154 VVVGHTLRKQGL-----VPLEKQELKANEVRLIYMSDSMPALASQIAGFIYAEFF 208
QY 243 NQOAVFKGIVPYITDELMKD-----TPKVFQVTQMAEKYNTYLAATKALIRAA 295
DB 209 NAAAEELKVGKYLRFQGDVWKNHACCVEFMHERDLTERAAMSQK-----VVDVVKAQ 261
QY 296 IYLDNMMNKKREATEMLA---OKQYGVADVETLAASNGTPEYKDDKRALPDFTFFR 352
DB 262 VETRA---HPOBAQLLSKSGNHYTPHSANVLT---VLAAPPDEBERYIAD----- 308
QY 353 HGASYPSSYSAWVYLTQLRRWGMINEFKPDNMYLDTAKV-YRPDIYLAAXELV----- 406
DB 309 -----RAIHHADM-----AKRIDQPRYPYATYBELVAKLA 341
QY 407 --AEGKAK-----AEDPPADTSIRPS 425
DB 342 TQVEGNAQFLHQLDPAFVARDLVDRFVKKS 372

RESULT 6
US-10-282-122A-44527
; Sequence 44527, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
```

```

; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 44527
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; FEATURE: MISC FEATURE
; NAME/KEY: (7)..(7)
; LOCATION: (7)..(7)
; OTHER INFORMATION: X=any amino acid
; US-10-282-122A-44527

Query Match
Best Local Similarity 9.3%; Score 227; DB 15; Length 257;
Matches 69; Conservative 52; Mismatches 122; Indels 50; Gaps 7;

QY 91 LDGSHMLAPLAPLAASVFGTKAD-----IEVPFMGNATVSNIEIWHQ-----KPMI 141
DB 1 LDAHGXSMPLPAAAMG-----ADQIGIALQTLVLSKRAFLISLOKLYQLALQESDA 56
QY 142 PLEGGKPVPIKADYLYKVEKYKAEKGFNNAMTFPAGSHNFKLRWYLAAGINPGYYS 201
DB 57 OTTAAQVIQYIQRD-----TLSLAVFKSHIHVCLREWLALA----- 95
QY 202 PPODISGOIGADALLSVTPPOMESTLEAGTIFGVCVGBPMNQAVFKGIVPYITDEEL 261
DB 96 -----DSRLAQTLKIKALPPPYVVALDNHVIDGFCVGBPMNQBELGLSKIVCSSODI 150
QY 262 WKDTEKVGVTQMAEKYNTYLAATKALIRAAIYLDNMMNKKREATEMLAQKQYGA 321
DB 151 IYVAVDKVLAVTQENABEGHPQTLVALLTTRIMQAQ--KELSNLKDPAPILKLL----- 200
QY 322 DVEVLAASNGTPEYKDDKRALPDFTFFRHHGASYPSSYSAWVYLTQLRRWG 374
DB 201 -VEFGIVRHCSEVHVDRKYWIMQNIYKYLKVENAPQEDPFHMLFQGWQKMG 252

RESULT 7
US-09-738-626-4906
; Sequence 4906, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEYA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
```

PRIOR APPLICATION NUMBER: JP 00/280988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: Patent ver. 3.0
 SEQ ID NO 4906
 LENGTH: 294
 TYPE: PRT
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-4906

Query Match 9.3%; Score 226; DB 9; Length 294;
 Best Local Similarity 23.6%; Pred. No. 4.8e-12;
 Matches 78; Conservative 56; Mismatches 152; Indels 44; Gaps 6;

QY 1 MKTIIRSSSKLLTLTSLAVGLTADVGAVGLEKED--LKEGFKLTDMAPLAV 57
 DB 1 MTHILFDSRFLQIGAFASLSTLGAARYVISTNNRPDNPPLTIGVPIAGSAPAI 60
 QY 58 AAEKGFEEDEGLFVQLEAQAQNVMDRVVNGELDSHMLAPPLASVGF-GTKADIEV 116
 DB 61 ADALGIFKGGVAVTLTKKYSQWSDLTWATAYATEQLDVAHMLSPMTVAINGVTNASRPTL 120
 QY 117 PSMGNGNAITVSNIMHQMKNIPLEGKPVHPITKADYLRKVERKYAEKGFPMKMT 176
 DB 121 SFTQNTGQAITLASKHYSSVNSADLKG-----MVLGIPFEYSV 160
 QY 177 PFGSHNIKLRWLAAGINPGYSPPODISQIGADALISVTPPQMPSTLEAGTIFGY 236
 DB 161 -----HALLIRDIYSNANP-----LADLELRKRPADMAQLTVEGIDGF 202
 QY 237 CVGEPNQAVFKIGVPIITBELMKDPEKVFEGVTQMAEKYPTVYLAATKALIRAI 296
 DB 203 ICGPFNERAISNGSGRIWLTQKQMDKHPCCAVMAKEMKAPHTAAQGVNALBESAS 262
 QY 297 WLDADNNKREKAIEMLAQKQYGVADVEYL 326
 DB 263 IL--SNPAQFDSSARTLSQEKYLNOPATLL 290

RESULT 8

US-10-282-122A-68363
 Sequence 68363, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:
 APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA 034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 PRIOR FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: Patent version 3.1
 SEQ ID NO 68363
 LENGTH: 372
 TYPE: PRT
 ORGANISM: Pseudomonas putida
 US-10-282-122A-68363

Query Match 9.1%; Score 222.5; DB 15; Length 372;
 Best Local Similarity 23.1%; Pred. No. 1.4e-11;
 Matches 90; Conservative 71; Mismatches 139; Indels 89; Gaps 17;

QY 42 LKEGFKLTDMAPLAVAAEKGFEEDEGLFVQLEA-QANMKVMDRVVNGELDSHMLAPA 100
 DB 17 VRIGVLPITDAPTLVAHNNGLFEAGIKAEKRVILRSMAQVIEAFISQGVNTHLSPM 76
 QY 101 PLASVGFETKADIEVPFSMGFNQNAITVSNIMHQMKNIPLEGKPVHPITADYLRKV 160
 DB 77 TWAR--YGSKVPAKVANNHVGSGLTVAADI-----SAVKQIGKTV----- 118
 QY 161 VEKYAEKGFPMKMTFPAGSHNIKLRWLAAGINPGYSPPODISQIGADAL-LSVT 219
 DB 119 -----AIFWYSIHNVVLAQQLNDGLTP--VSKP--ANAQLAANEVNLVL 161
 QY 220 PPPMPSTLEAGTIFGYCVGEPNQAVFKIGVPIITBELMKD-----TPEKVF 272
 DB 162 PPEDMPALASAKRIAGTYIAEPNALAENIKVGRVQFTGDVWRNACCVFPHENHLN 221
 QY 273 TKQMAEKYPTVYLAATKALIRAIWLDADNNKREKAIEMLAQ--KQYGVADVEYLAAS 329
 DB 222 RPEWSQK-----VYNALVKAQW---TRDRTEAALLSPAGPKVYPPHBAVLTKV 270
 QY 330 M-----NGTFEYKDKRALPDNTPFRIGASIPSSAVMYLTOLRMGMIN 377
 DB 271 LAPAEDRAGYIASGAIHQWDEKRI-DFQPY-----PPSYTEB---LVKALKTTLLI- 320
 QY 378 EFKDNMYLDTAKNVRPDIYLAARELY 406
 DB 321 --EGDNTFISGLDPAY-----AARDLV 340

RESULT 9

US-10-282-122A-49957
 Sequence 49957, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:
 APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA 034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 PRIOR FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PAM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 49957
LENGTH: 391
TYPE: PRT
ORGANISM: Burkholderia fungorum
US-10-282-122A-49957

Query Match 9.0%; Score 220.5; DB 15; Length 391;
Best Local Similarity 24.4%; Pred. No. 2,4e-11;
Matches 94; Conservative 63; Mismatches 136; Indels 93; Gaps 16;

9 SKKLLTTSASLAWGLTAPDVGAIVKLEKED---LKEGFIKLTMAPLVAAEKGF 64
2 SREMLKLSMTYVAG--AGPLIGALNAPADPNAPVIGIPIITDPTLVVANNNGF 59
65 EDEGFEVQLEA-QANMKVMDRVVNGELDSHMLAPAPLAASVGFCTKADIEVPSMGN 123
60 EAAGLAVEKPTLLRSWAQLVEAFISGVVWVHLLAPMTIAR--YGSQAPAKVVMNNHN 117
124 GNAITVSENIHQMKPNIELEGKRVHPKADYLKVEKTYAEKGKPFMMANTFPAGSIN 183
118 GSALTIVSPDI-----GKVADLGKTV-----APFWYSIIN 148
184 IKLRFWLAAGINP-----GYSPDIDISGIGADALLSVTPPQWSTLEAGTFGCV 238
149 VVVQMLRQGLVPLVKDGEIKP-----NEKRLVMAPSDIPPLASRQJAGFTV 199
239 GEPNQOAVFKSIGVPIITDEELMKD-----TEKVFQVTKQNAEKYPTNYLAVTAL 291
200 AEPFNAABEGIKIGKVLRTGPDVWKNHACCVFPMHERDITERAANSQK-----VDSV 252
232 IRAATWLDADNNKRNKEAIEMLA---QKQY---VGADVEVLAASNGTFEYKDKRAL 344
253 VQAQMLTRA---HFOEAQQLSKTGQNMHTPTTAVALTEVLAPAGDGRYIAD--RAI 306
345 -----PDFTFFRHGASVPSYS 361
307 IHADWAKRIDQPY-----PYPAIT 327

RESULT 10
US-10-627-476-384
Sequence 384, Application US/10627476
Publication No. US20040030116A1
GENERAL INFORMATION:
APPLICANT: Pompejus, Mark
APPLICANT: Kroger, Burkhard
APPLICANT: Schoder, Hartwig
APPLICANT: Zelder, Oskar
APPLICANT: Habersauer, Gregor
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
TITLE OF INVENTION: TRANSPORT
FILE REFERENCE: BGI-125CPGN

CURRENT APPLICATION NUMBER: US/10/627,476
CURRENT FILING DATE: 2003-07-25
PRIOR APPLICATION NUMBER: 09/602,787
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: USSN 60/141031
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: DE 19931454.3
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931478.0
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931563.9
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19932122.1
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932124.8
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932125.6
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932128.0
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932180.9
PRIOR FILING DATE: 1999-07-09
Remaining Prior Application data removed - See File Wrapper or PAM.
NUMBER OF SEQ ID NOS: 678
SEQ ID NO 384
LENGTH: 275
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-10-627-476-384

Query Match 9.0%; Score 219.5; DB 15; Length 275;
Best Local Similarity 23.1%; Pred. No. 1.7e-11;
Matches 66; Conservative 52; Mismatches 127; Indels 41; Gaps 5;

42 LKFGFIKLTMAPLVAAEKGFEEDEGLFVQLEAQNKKVMDRVVNGELDSHMLAPAP 101
26 LTIGVYPIAGSPALADALGIFKKGAVNYTLKXSGMSDMLTATBEDLVANHLSPT 85
102 LAASVGF-GTKADIEVPSMGNALIVSENIHQMKPNIELEGKRVHPKADYLKPV 160
86 VAINAGVTNASRPTLSFTQNTNGQAITLASKHYGSVNSADLKG----- 130
161 VEKTYAEKGKPFMMANTFPAGSHNITLRYWLAAGINPGYSPDIDISGIGADALLSVTP 220
131 ---WLDGIPFYSV-----HALLRDYLVSNAVDP-----IADLEIRLR 167
221 PPMQSTLEAGTIFPYCVGEPNQOAVFKSIGVPIITDEELMKDTEPEKVFQVTKQNAEKY 280
168 PADWQAQLTBEDIFIGRPFNERAISNGSRITWLTQKMDKRPCCAVMAKEMKAEH 227
281 PNTYLAVTKALIRAIWLDADNNKRNKEAIEMLAQKQYVGADEVYL 326
228 PTAQGVNALAEASAIL--SNPAQFDSASRTLSQEKYINQPATLL 271

RESULT 11
US-10-282-122A-48182
Sequence 48182, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48182
LENGTH: 391
TYPE: PRT
ORGANISM: Burkholderia cepacia
US-10-282-122A-48182

Query Match 8.9%; Score 217.5; DB 15; Length 391;
Best Local Similarity 23.3%; Pred. No. 4.5e-11;
Matches 95; Conservative 67; Mismatches 130; Indels 115; Gaps 16;
9 SKKLLTSLASLAWG---LTTA-----PDVGAAGKLEKEDLKFGFIKLTMDAPLAV 57
2 SREBWKLSALPTATGAAPLISMANAHAAEPDA-----PVRIGYLPITDAADLV 52
58 AAKGPEDEGLTV-QLEAQAAMKVMDRVNGBLDSHMLAPPLAASGFGKADIEV 116
53 AHNNGFASGLTVYEQKLRSWAQVLEAFISGVNVNHLSPWTLMAR--YGSRAAKV 110
117 PFMGNGNAITVSNBIHQMKENIPLEGKPVHPIKADYLKPVKEKYKAEKPFNNAMT 176
111 VAMNHNNGSLTTPADPDI-----DSLRDLGKTV-----AVP 141
177 PPAASHNIKLRVWLAAGINPGYVSPQDISGQIGADALLSVTPPQMPSTLEAGTIFGY 236
142 FWTSHNVVLQDMLRGGLVPLKRGSTPAANEVN---LVVMAPSDMLPALAARQIAGY 197
237 CVEPNNQAVFVGIGIVPVTTDEBLMKDTPEKFGVTQKQ-----WAEKPYTYLAATK 289
198 IYAEFPATATELWKIGILFTFGVWKNHACVVFMEHODLTPORPAWSOK-----VVN 250
290 ALIRAAIWLADNNKKEAIEMLAQKQYGVADVEVLAASNGCFEEXD-----DKRAL 344
251 AIVKAQDMA---RSHQETAOQL-----SKDGHKHSPTLASLDRVLV 291
345 PD---FNTFPRHGA-----SYSPYSASVWYLTQLRR 372
292 PSASLADTVRASGIRHADWHAKRIDPQYRPFPSYTEAL--VQRLKR 336

RESULT 12
US-10-282-122A-72603
Sequence 72603, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 72603
LENGTH: 342
TYPE: PRT
ORGANISM: Streptococcus mutans
US-10-282-122A-72603

Query Match 5.9%; Score 144; DB 15; Length 342;
Best Local Similarity 21.4%; Pred. No. 0.00022;
Matches 83; Conservative 59; Mismatches 145; Indels 100; Gaps 15;
12 LLLTSLASLAWGTLTPDVGAAGKLEKEDLKFGFIKLTMD--APLAAAEKGFEEGL 69
15 ILPTISVSLVACSKSSTSTSKNKNYASEKDSYILKISNSDLGAPQDIAIEKGFDDVGL 74
70 FVQLEAQAAMKV---MDRVNGBLDSHMLAPA---PLAASVGFTKADIEVP 118
75 -----KYKVAKIGODTSLNLDNAAGKIDASNSLMASTIOPLA----- 111
119 SMGFGNNAITVSNBIHQMKENIPLEGKPVHPIKADYLKPVKEKYKAEKPFNNAMT 178
112 -----NGAKKITITGL-HTGGLQILTKNGK-----IKSAABLKGGKIGTVAV 152
179 AGSHNIKLRVWLAAGINPGYVSPQDISGQIGADALLSTPPQMPSTLEAGTIFGYCV 238
153 AGSPALFAKVLAKSGIK-----VSDKKG-DVSFVYQSDQLQOVLDKGEVDIAL 202
239 GEPNNQAVFVGIGIVPVTTDEBLMKDTPEK---VFGVTKQMAEKYPTYLAATKALIRAA 295
203 GDP-DTEVLKKQYGFETLANSSDTKGFKNMECCVAYVSNITVKKHPVAAKTYTLMQKAA 261
296 IWLADNNKKEAIEMLAQKQYGVADVEVLAASNG--TFEYKEDKRALPDFTFPRHG 354
262 NWV-----QKKEKEFTVDIQLNQNVVAGSKDNLSTLSNYTK----- 298
355 ASYPSYSSAV---WYLTQLRRWGMIN 377
299 ---PSYSGAYDSPTVASDLRKIGIIS 322

```
RESULT 13
US-10-282-122A-47676
; Sequence 47676, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Mang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forey, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 47676
; LENGTH: 192
; TYPE: PRM
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47676

Query Match      5.7%; Score 140; DB 15; Length 192;
Best Local Similarity 26.8%; Pred. No. 0.00021;
Matches 42; Conservative 32; Mismatches 51; Indels 32; Gaps 4;

QY 42 LKFGFKLTDMAPLVAARKGFPEDEGLTV-OLEAQAANKVMDRVVNGELDGSHTLADPA 100
DB 37 VIGIYLPITDAAPLVAHNNHGYFASGLTVEQPKLRSWAOLVEAFLSGQVNVVHLSFM 96
QY 101 PLAAAGVGTCKDIEVPSMGFNGNAITVSNELWQMKNNIPLEGKGPVHPKADVLKRV 160
DB 97 TLMAR-YGSRAPAKVANNHNGSGLTVAPDI-----DSLRLDGGKTV----- 138
QY 161 VEKYAKGKPFMMAMTFPAGSHNIKRLWYLAAGINP 197
DB 139 -----AIPFWYSIHNVTVQDMLREGGLVP 162

RESULT 14
US-10-156-761-8511
; Sequence 8511, Application US/10156761
; Publication No. US20030119018A1
```

```
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8511
; LENGTH: 883
; TYPE: PRM
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8511

Query Match      4.8%; Score 116.5; DB 14; Length 883;
Best Local Similarity 21.3%; Pred. No. 0.32;
Matches 114; Conservative 62; Mismatches 147; Indels 213; Gaps 35;

QY 6 RSSSKKLLTSLASLAWGTLTAPDVA-----VGKLEKEDLKFGFKLTDMAP-----L 55
DB 15 RSLGRSLALTMTAVLGAAGLALSPASADSGLPTTADPTFAHGLKGEY--YTQSAFGAPDF 72
QY 56 AVAAKGFPEDEGL-FVOLEAQAANKVMDRVVNGELDGSHTLAPPLAASVGFGRKADI 114
DB 73 ASIKATGF--DGLDPLDSLSESR-----LSAATGRADVSV 105
QY 115 E-----VP-----FSM-GFGNNAITVSENI-----W--HQMKNNIPLEGKGPVHP 151
DB 106 RWTGEVPEKSGPTTPTSTGDNGLFRLWVQKLTIDHWUDWIREQSGEVDLTAGR-AVD 164
QY 152 IKADYLKPVVEKYKAKGKPFMMAMTFPAGSHNIKRLWYLAAGINGYVSPQ----- 204
DB 165 IKVEYEHF-----CGSNLHLR-WTPPGSKS---AVPQSAFRLPD 201
QY 205 --DISQIGADAL-----ISVTPPPQMP-STLEAGTTFGYC 237
DB 202 GPDYNGALSATVLDGGRVRLDPPALASAPATVLDHLTVTSGKKWPLSSVKT----- 254
QY 238 VGEPMNQAVFGKIGVPIVITDE-----ELMKD-----TPEKVPGYTKOMARK 279
DB 255 --DPSDARALLVGLGEPVGNKAGTRAKGVNVTYDGKALADTSGTPVAF-----WSSG 307
QY 280 YPN--TYLAVTALIRAAITMDADNNKKEAIEMLAOKOYVADVEVLA-ASMNGTFEY 336
DB 308 -FNKSTYELARTK-----WADQVGPKN-----ALPEYRPPQLTRKAMQNLNGTWP 351
QY 337 -----EKODGRALPDFNTFFRHGASYP-----SYSAWYTLTLBRKMGINERK 380
DB 352 AGAKAGEEPRTKTLAE-----RIIVPYVESQLSGIERHEDRMWY-----RTFTV----- 398
QY 381 PDNWTYLDTAKVY-----YRPDIYLAAKELVAGKAKADFPMDT--STKPSQ 426
DB 399 PADMKVGHGKRLRLNFGAVDMQSEVYVNGTK--VAEHKGGYRKFSADVTDALPKGR 452

RESULT 15
US-10-156-761-9969
; Sequence 9969, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
```

APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9969
LENGTH: 353
TYPE: PR1
ORGANISM: Streptomyces avermitilis
US-10-156-761-9969

Query Match 4.6%; Score 111.5; DB 14; Length 353;
Best Local Similarity 18.4%; Pred. No. 0.23;
Matches 71; Conservative 55; Mismatches 161; Indels 99; Gaps 12;

QY 5 IRSSSKLLTLTASASLAVWGLT-IAPDVGA-----GKLEKEDLKFGPIKLTDMA 53
DB 1 MRETAARYTALAAGLALSSLTACANDASTASSGSGSKDGGVKYKIMWGGIDKVIYL 60
QY 54 PLAVAAEKGFPEDEGLFVQLEAOANWKVMDRVVNGELDGSHMLAPAPLAASVGF----- 108
DB 61 PAULTQRLGTFDEAGLDVELSEPAVGQAEALVSGVQ-----AVGFYDHTL 109
QY 109 -----GTKADIEVPFSGKFNAGNAITVSNELIWHQMKPNIPLGGKPVHPIKADYIKPVER 163
DB 110 DLQTKGKDVSVOFSHA-PGEVEIVSNR-----HADITSP 145
QY 164 YKAEGRPFNNMAMPFPGASHIKLRMYMLAAGINPGYSPPODISGOIGADALLSYTPPQ 223
DB 146 KDFKGR-KLGITGLSSSTDFLTXYLAVKXGVSSEFSP-----VAVGAGPT 190
QY 224 MPSTLEAGTIFGYCVGEPNNQOAVFKGIGVPI-----TDELMKDTPEKVGVTQKW 276
DB 191 FISALQKGSIDGGMTTDPVTANILAKNLGKVLIDMTFRGSKKALGGPYFSSSLYMQTEW 250
QY 277 AEKYPNTYLAVTKALIRAAIWLADANNKRNKEAIEMLAQK---QYVGADVEVLAASNGT 333
DB 251 VNSHKOTVQKLANAFVYTKLMGT-----HTASQIAEKMPADYSGGNKTLVAGAIKTT 303
QY 334 FEYKXD-----KRALPDFN 348
DB 304 LPMFTDDGVMPKDGPTVERVLKAFN 329

Search completed: May 18, 2005, 17:49:23
Job time: 147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2005, 17:36:01 ; Search time 57 Seconds
(without alignments)
783.238 Million cell updates/sec

Title: US-10-689-200-2

Perfect score: 2443
Sequence: 1 MTTIRSSSKLLTLTSLASL.....AKFALGKQQTAVAGKVD 464

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	26.6	670	2 S77387	nitrate transport
2	627.5	25.7	670	2 AC3037	hypothetical prote
3	627.5	25.7	471	2 H98248	nitrate-binding pr
4	616.5	25.2	441	2 E95928	probable nitrate t
5	611.5	25.0	431	2 AD3037	hypothetical prote
6	611.5	25.0	431	2 G98248	hypothetical prote
7	610.5	25.0	440	2 AG1882	nitrate transport
8	587	24.0	657	2 A11882	nitrate transport
9	576	23.6	430	2 C95300	Nrta-type periplas
10	572	23.4	443	2 JQ2134	nitrate transport
11	549.5	22.5	430	2 G87324	hypothetical prote
12	545.5	22.3	442	2 S58738	nitrate-binding pr
13	538.5	22.0	667	2 AH2165	bicarbonate transp
14	534.5	21.9	667	2 S75959	nitrate transport
15	525	21.5	446	2 S77389	nitrate transport
16	502.5	20.6	458	2 AF2165	bicarbonate transp
17	494	20.2	402	2 F83422	conserved hypochet
18	476	19.5	659	2 S30893	nitric protein - Syn
19	473.5	19.4	452	2 S75956	nitrate transport
20	470.5	19.3	427	2 AF2222	nitrate-binding pr
21	444	18.2	450	2 A44751	carotenoid-binding
22	398.5	16.3	392	2 S52248	nitrate and nitrit
23	390.5	16.0	426	2 AF2663	hypothetical prote
24	390.5	16.0	426	2 B97445	nitrate transport
25	355	14.5	625	2 S58789	ATP binding protei
26	249	10.2	225	2 B55581	naef protein - Kle
27	240	9.8	399	2 AB3354	hypothetical prote
28	125	5.1	337	2 B75370	probable ABC trans
29	124.5	5.1	333	2 AB3080	ABC transporter, s

30	124.5	5.1	333	2 G98206	hypothetical prote
31	116.5	4.8	468	2 AB3359	hypothetical prote
32	114	4.7	325	2 H75414	probable ABC trans
33	114	4.7	336	2 AB3801	ABC transporter (s
34	111	4.5	310	2 AC3238	hypothetical prote
35	109.5	4.5	300	2 H69260	thiamin biosynthes
36	106.5	4.4	675	2 D48492	kpac protein - Bbc
37	106	4.3	483	2 G86811	sugar ABC transport
38	106	4.3	1273	2 T34558	hypothetical prote
39	105.5	4.3	339	2 E95962	probable taurine u
40	105.5	4.3	566	2 A40589	cellulase (EC 3.2.
41	104.5	4.3	1135	2 T14659	DNA polymerase III
42	104.5	4.3	1174	2 T15021	probable DNA polym
43	103	4.2	897	2 AC2001	hypothetical prote
44	102	4.2	301	2 B95947	phosphate uptake A
45	102	4.2	1291	2 I40631	non- proteolytic bo

ALIGNMENTS

RESULT 1

S77387 nitrate transport protein C-2 - Synechocystis sp. (strain PCC 6803)

N/Alternate names: protein s11452

C/Species: Synechocystis sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 16-Aug-2004

C/Accession: S77387

R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimp, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

A/Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S77387

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-670 <KAN>

A/Cross-references: UNIPROT:P73450; EMBL:D90906; GB:AB001339; NID:G1652492; PIDN:BAA1749(C

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C/Genetics:

A/Gene: nrtC-2

C/Superfamily: ATP-binding cassette homology

C/Keywords: ATP; nucleotide binding; P-loop

P/25-215/Domain: ATP-binding cassette homology <ABC>

F/42-49/Region: nucleotide-binding motif A (P-loop)

Query Match 26.6%; Score 651; DB 2; Length 670;

Best Local Similarity 36.3%; Pred. No. 9, 7e-40;

Matches 158; Conservative 68; Mismatches 153; Indels 56; Gaps 13;

QY	28	ADPVGAVGK-----LEKEDLKFGFKLTMTMALVAAGKFFDEBL-FVQLEAOA 77
DB	260	AKKVGAVSOFPAEMGANGLEKINLIDGFTPLTDCAPLVAKXGFGPKHLEOVNLKPEP 319
QY	78	NKVVMDRVYNGELDGSMLAPAPLAASVFGTKADIEV--PSMGNGAIVTSNEIWH 135
DB	320	SWQAIADGIRERRLDAGQVAGMPALTLTGMGSKTLPVMTAMVBSNGAIIISKFFAE 379
QY	136	QMKNPILBEGKGVHPIKADYLLKPVVEKYKAEGKPRFMAMTPPAGSHNIRRYWLAAGI 195
DB	380	-----AGVKLTLEBLR-----LKLAEETPPQVSTLGVHVAHQNLRLRWLASGSI 424
QY	196	NPGYISPPQDISQIQADALLSVTPPPQPESTLBAGTIRGYCYGEPNQOAVKGIQV 255
DB	425	D-----PDQDIN-----LKRLLPPOVWSNLEAGNIDGFCVGBEPWNSYAVQNGYVI 471
QY	256	ITDEELMKDPEVPEVPTVQMAEKYQNTYLAATVLAIRAIWLDADNNKRAIEMLAQ 315
DB	472	ATLDLIMNGHPEVYLVGLRREWVVKIPATHLALVKALLBACEY--CDRRHROEIIYLLAL 529
QY	316	KQYVGVADVEVLAASNGTPEYERD-DKRALPDENTEFRHGASYSYSSAVVYLTOLRWG 374

Db 530 POUYGVSTEVISGFLTEYVQDAEEMLLDPNOGTVKSNTPSSEGMILTOLARK 569
QY 375 MINEFPMWYLDTAQNVRPDIYLAALAEVAGKAKADEFPADTSIKPSQNFIDKVP 434
Db 590 YID--PPKMW-VIIRVRPDLFGCAEACRL---GMPDLGDHNNVSL-----PDGMV 636
QY 435 PDANKENDYAKFAI 449
Db 637 FTNPDELGIKRTI 651

RESULT 2

AC03037
hypothetical protein nrtA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: AC03037
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AC03037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-430 <RUR>
A:Cross-references: UNIPROT:Q8U925; GB:AE008689; PIDN:AAU44713.1; PID:g17742344; GSPDB:G
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: nrtA
A:Map position: linear chromosome

Query Match 25.7%; Score 627.5; DB 2; Length 430;
Best Local Similarity 35.5%; Pred. No.2.7e-38;
Matches 166; Conservative 70; Mismatches 171; Indels 61; Gaps 19;

QY 1 MKTIIRSS-SKLLD--TTSASLAWGLTIAPD--VGAVGKLEKEDLKRGFIKLTMAPL 55
Db 1 MKKIFSGTVSRRTILKTTATTAALVTAVRTAPSGAPAAVABEVKAKLGFALTDAAPL 60
QY 56 AVAAEKGFPEDEGL-FVQLEAOANMKVNDRVNG---ELDSHMLAPAPLAASVFGT 110
Db 61 IIAAEKGLPAKKGMPVEVLKQASWGATRDNLVLGGASNGIDGAHILTPMYLMTGKXT 120
QY 111 KADIEVPS--MGFNCAITVSNIEIWMOKNPILBEGGKPVHPKADYLYKPVVEKYA 166
Db 121 QNNVPVPMITLARIINDSOGISVAKE---YAETGVOLDASK-----LKAAPKKA 168
QY 167 ECKPFNMAMTPPAGSHNIIKLRYWLAAGINPGVYSPPOISQIGADALLSVTPPOMS 226
Db 169 DKEIKAAKMTFPGTHDLWIRYLAAGID---PDKVS-----TIIVPPQWVA 215
QY 227 TLEAGTIFGVCGEPMNOQAVFKIGVPYITDELMKDPPEKFGVTQMAEKYPTTYLA 286
Db 216 NKKVGMDFVCVGEPMNEQLVNOGIGFTACTTGELMKGHPKALGRRADWVEKNPATKA 275
QY 287 VTKALIRAIWLDADNNKRAIEMLAQOYVADVEVLAASNGTFEYKDDKRALPD 346
Db 276 LMAVMAEQM--CEMANKEEMSTILGKQWPNVPKQVGLRLKGNINY--GNGRVLE- 330
QY 347 FNT-----FRRHGAQSPYSASVWYLTQLRMGMINFEKPDWYLDTAQNVRPDIYLA 401
Db 331 -NTGLOMKFWQDHASVPPSHDSWFTTENTIRNG--KFAPDVDVAKLVAKVREDIWRRA 386
QY 402 AKELVAGKAKADEFPADTSIKPSQNFIDKVPDANKENDYAKFAI 449
Db 387 AKDL---GVA---DLPASTS-RGKETFPDGKV-FPEPNSAVLESISI 426

RESULT 3

H98248
nitrate-binding protein nrtA precursor, periplasmic [imported] - Agrobacterium tumefaciens
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C/Accession: H98248
R/Godner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Queroljo, B.; Goldman,
A.; Liu, P.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98248
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-471 <RUR>
A:Cross-references: UNIPROT:Q8U925; GB:AE007870; PIDN:AAK89514.1; PID:g15159390; GSPDB:G
A:Genetics:
A:Gene: AGR L 1886
A:Map position: linear chromosome

Query Match 25.7%; Score 627.5; DB 2; Length 471;
Best Local Similarity 35.5%; Pred. No.3.1e-38;
Matches 166; Conservative 70; Mismatches 171; Indels 61; Gaps 19;

QY 1 MKTIIRSS-SKLLD--TTSASLAWGLTIAPD--VGAVGKLEKEDLKRGFIKLTMAPL 55
Db 42 MKKIFSGTVSRRTILKTTATTAALVTAVRTAPSGAPAAVABEVKAKLGFALTDAAPL 101
QY 56 AVAAEKGFPEDEGL-FVQLEAOANMKVNDRVNG---ELDSHMLAPAPLAASVFGT 110
Db 102 IIAAEKGLPAKKGMPVEVLKQASWGATRDNLVLGGASNGIDGAHILTPMYLMTGKXT 161
QY 111 KADIEVPS--MGFNCAITVSNIEIWMOKNPILBEGGKPVHPKADYLYKPVVEKYA 166
Db 162 QNNVPVPMITLARIINDSOGISVAKE---YAETGVOLDASK-----LKAAPKKA 209
QY 167 ECKPFNMAMTPPAGSHNIIKLRYWLAAGINPGVYSPPOISQIGADALLSVTPPOMS 226
Db 210 DKEIKAAKMTFPGTHDLWIRYLAAGID---PDKVS-----TIIVPPQWVA 256
QY 227 TLEAGTIFGVCGEPMNOQAVFKIGVPYITDELMKDPPEKFGVTQMAEKYPTTYLA 286
Db 257 NKKVGMDFVCVGEPMNEQLVNOGIGFTACTTGELMKGHPKALGRRADWVEKNPATKA 316
QY 287 VTKALIRAIWLDADNNKRAIEMLAQOYVADVEVLAASNGTFEYKDDKRALPD 346
Db 317 LMAVMAEQM--CEMANKEEMSTILGKQWPNVPKQVGLRLKGNINY--GNGRVLE- 371
QY 347 FNT-----FRRHGAQSPYSASVWYLTQLRMGMINFEKPDWYLDTAQNVRPDIYLA 401
Db 372 -NTGLOMKFWQDHASVPPSHDSWFTTENTIRNG--KFAPDVDVAKLVAKVREDIWRRA 427
QY 402 AKELVAGKAKADEFPADTSIKPSQNFIDKVPDANKENDYAKFAI 449
Db 428 AKDL---GVA---DLPASTS-RGKETFPDGKV-FPEPNSAVLESISI 467

RESULT 4

E95928
probable nitrate transport protein nrtA [imported] - Sinorhizobium meliloti (strain 1021)

C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C/Accession: E95928
R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernant
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95928
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <RUR>
A:Cross-references: UNIPROT:Q92VK8; GB:AL591985; PIDN:CAC49093.1; PID:g15140578; GSPDB:G
A:Experimental source: strain 1021, megaplasmid pSymB

A:Gene: nrCA, SMD21114
A:Genome: plasmid
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, A.; Rebuffat, P.; Vandenhof, M.; Votrholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: nrCA, SMD21114
A:Genome: plasmid

Query Match	25.2%;	Score 616.5;	DB 2;	Length 441;
Best Local Similarity	33.8%;	Pred. No. 1.8e-37;		
Matches 153;	Conservative 74;	Mismatches 182;	Indels 43;	Gaps 12

```
QY      SLAWGTLTAPDVAVGVLEKEDLKFGIKITDMAPLATAAEKGPFEEDEGFVLOEAQN 78
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      NLSAGGLPAP--GSIRHESRTVRAGTILPVDAANVLAALAEFFQAQREGITLIELVDVS 76
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY      WKVMVDRVANGELDGSMLAPAPLAASVGFTKADIEV-PSWGFNGNAITSNEIWHQM 137
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      WANVRDLRAFRQFVAHVHLSPMPAAMLGISNNSPTIAPSLRGGAIIITLSTRLGLM 136
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY      KPNILPLEGGKVHHIKADYLKEPVVEKYKAEEK-PENMAMTPEGASHNKLRMYTLAAGIN 196
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      QOAGCLGGGEBALN-NKALLAAVIRISAGAACKPLTTIGVTYPSSHTEFEFYTLAAGID 195
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY      PGYYSPODISGOIGADALLSVTPPPOMPSLTBAGTIIFYGCVCSEPMNQOAIFKGIQVPVI 256
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      P-----DROVKLVVVPPEMTSDALAAGAIIDGFCVGA PMNVVASERGVRI VA 242
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY      TDEELMKOTPEKVIVGTWKQNAEKT ENTYLAATKLLIRAITLWDLDNNKNRKEALEMTLAK 316
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      TKODIPWSAPEKPIGVMRPEWAENARDTVSRVLVALLDRAAR--SDEBANRGRLEVALEE 300
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY      QYVADVREVLAAASNUGTFEYEKD-DKRALPDENFFRRHGASPYSSAVWTJLTOBRGM 375
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      RHVAAPADITIRRVLAGEFALDPBGNRKRVENEVLYFHAGFANYPRPSQALMTYTSQMWTKQ 360
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY      INEFKPNNWYLTDAKNYRPDIYLAAKELVABEGAKAKADEPADTISI----KPSQNFPI 430
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      TSLSRQ---RVDALISAIRPDIYREAL-----GK---DALPADADAQRLEGATDGDRFM 408
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY      DKXVPEDANKRPNDYLAKFALGLKQGYTAGSKV 462
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      DGHFVDPGRIEDYIKSF-----GAPSVVNSRY 435
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
```

RESULT 5
AD3037
hypoetical protein nrcC [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AD3037
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monk, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McTiell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
eter, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:21608550; PMID:11743193
A:Accession: AD3037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-431 <KUR>
A:Cross-references: UNIPROT:O8U924; GB:AB006689; PIDDN:AAL44714.1; PID:g17742345; GSPDB:C
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: nrcC
A:Map position: linear chromosome

Query Match	25.0%;	Score 611.5;	DB 2;	Length 431;
Best Local Similarity	35.5%;	Pred. No. 4.1e-37;		
Matches 151; Conservative	69;	Mismatches 166;	Indels 39;	Gaps 12;

```

QY 34 UGKLEKEDDKPFPIKLTIDWAPLAVAAEKCFEEDGCFVOLEAQAMMYMDRUVNGEILDG 93
Db 24 VGSDBQKILIRAGITPLVDASVLIAABEFGFADBEGLILDVKOVSMANVRDRLFRQDPI 83
QY 94 SHMLAPAPLAASVGFCTKADIEV-PSMGFNGNAITVSNBIMHQMKNPIPL-EGGKPVHP 151
Db 84 AHMLSPMPAPASMLGIGSNPSPTITPSTLRGCAVITLSTRLPARMKALTGLSETAGALEN 143
QY 152 IKADYLYKPVVEKXKAEKG-PFNMAATFPAGCSNHIKLRITMLAAGINGYISPPDISQI 210
Db 144 ARA--LKIYLLDMRARGEARPTLGMTYPPSSSHYERRIYMLAAGIHDPH----- 190
QY 211 GADALLSVTPPROMPSTLAGEITFGYCVBEPNQOAVFKSIGVYIIDEELMKDTPKEKVF 270
Db 191 --DYKLVAVPPMTSALNAGAIDGFCVGAPMNIVAAERBGVRVAAKQDIMPAPERYI 248
QY 271 GVTQMAEKYPTNYLAVTKALIRAAIMLD-ADNNKRRKEALIEMLAQOYVAGADVEVLAAS 329
Db 249 GMRPEMAESQOQETVGRLLTALDAAASWCDLADNHDALSGA---LADPRYIGAPOSIIRRV 305
QY 330 MNGTEYE- XDDKRALPDNTPFPRHGAISPYSSASAWYLTQLARRMGKINERKPDNWTYDT 388
Db 306 LAGEFSISQGNRRVIEKFTFPHGDHANYPROSQSLMYSQMLRWG---QALSETGVNA 362
QY 369 AKNVVRPDYLYLAARELVAAEKAKADFPADTSIK-----PSQNFIDKVPDPDAKNPNDYL 444
Db 363 ALSAVRPDIYRA---LQDGKA-----PGDADIRIEGDEGBRFDVGVFDPDADIAGYV 413

```

445 AKFAI 449
 Oy ||:
 Db 414 NSFAV 418

 RESULT 6
 G98248
 Hypothetical protein AGR_L1881GL [imported] - Agrobacterium tumefaciens (strain C58, Cerevisiae)
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C:Accession: G98248
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
 A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-431 <KUR>
A:CROSS-references: UNIPROT:Q8U9Z4; GB:AE007870; PDB:AAK69513.1; PID:g15159389; GSPDB:G
A:Gene: AGR_L_1881g1
A:Map position: linear chromosome

Query Match 25.0%; Score 611.5; DB 2; Length 431;
Best Local Similarity 35.5%; Pred. No. 4,1e-37;
Matches 151; Conservative 69; Mismatches 166; Indels 39; Gaps 12;

CY 34 VGEKEEDLKFGFIKTDMAPLVAALAKGFEEDEGLFVQLERQAAMKYVMNDRVNGBELDG 93
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 24 VGSDRÖKITLRAFIPLVDASVLIAALEFGFADEEGILLDLVKDVSMAVNDRILAFROFDI 83

CY 94 SHTLAAPLPAASVGCEFTKADIEV-PFSMGFNCAITVTSNEIMQHCKNPIL-ECKGFVHP 151
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 84 AHMSPMVPVASMLTGLGSNPSPTITPSSLRCGNAILTLSTRLPARMKALTGLSETAGALEN 143

CY 152 IKADYLTKVEVEKKYKAEGK-PFMNAMTFPAGSHNIKIARYTLAAGINNGYSSPRODISGOI 210
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 144 AAA--LKLVLDMRRARGEARPTLLGMTYPPSSHAYERRRYWTLAAGGIHDPD----- 190

QY 211 GADALLSVPPQMPSTLEAGTIFGVCVGEPMNQAVFKGIVPVITDEELMKDTPKVP 270
 DB 191 --DVKLVPPEPTSDALAGIDGFCVGPVVAERKVGRIVAKKDLMSAPKVI 248
 QY 271 GVTQNAEKYENTYLAATKALIRAAIWL-D-ADNNKRNKEAIEMLAQOYVGADEVVLAAS 329
 DB 249 GMRPENAESQOETVGRLLTLMDAAASWCLADNHDLSGA---LADPRYICAPQSIIRRV 305
 QY 330 MNGTFEYE-KODKRALPDENVTPFRHGASYSYSAWVYLTQLRRWGMINEFKPDNYLDT 368
 DB 306 LAGEFIDSQGNRRVIEKYFTFHGDHANYPROSQSLIYQMIWRWG---QAEISGTVNA 362
 QY 389 AKVVPDYIYLAAKELVAEGKAKADEFPADTSIK---PSQNFIDKVPEDANKENDYL 444
 DB 363 ALSAIPDIYRAA---LGDGKA-----PDADIRIBGDEGRFVDEVPADIAGYV 413
 QY 445 AKPAI 449
 DB 414 NSPAV 418

RESULT 7

AG1882
 nitrate transport nitrate-binding protein nrtA [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C/Accession: AG1882
 R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Status: preliminary
 A:Accession: AG1882
 A:Molecule type: DNA
 A:Residues: 1-440 <KUR>
 A:Cross-references: UNIPROT:Q44292; GB:BA000019; PIDD:BA072566.1; PID:g17129954; GSPDB:G
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: nrtA

Query Match 25.0%; Score 610.5; DB 2; Length 440;
 Best Local Similarity 33.4%; Pred. No. 36-37;
 Matches 158; Conservative 71; Mismatches 173; Indels 71; Gaps 17;
 QY 8 SSKKLLLT-----LSASLAVWGLTI-----APDVGAVG--KLEKEDLKF 44
 DB 5 SRKKELFTTGAAAASILVHGCTSNQSOSATTEQAPSAAPAAVVAAPKVEITKAL 64
 QY 45 GRITLTMAPLVAAEKGFEEDEGLF-VQLEAQANKVVMDRV---VNGELDGSHTLAP 99
 DB 65 GEIPLTDAPLIIAEKGFPAKGMTDIVIKQKSPVTRDNLKIGSSGGIDGAHILSP 124
 QY 100 APLAASVGGTADIEVPPSMGNGNAITVSNIEIHQMKPNIPLEGKRVHPKADYLP 159
 DB 125 MFLYMTI--NDKVPMYTLARLNTNGOASIVAEKF--KELAVLE-----SKSLKD 170
 QY 160 VVEKYAEKGFPMNMTFPAGSHNITKLRYWLAAGINPGYSPPOISQIGADALLSVT 219
 DB 171 AAIKAKADKKALKMGITFGTHDLMRKRYWLAAGINP-----DQDVLEAV 217
 QY 220 PPPQMPSTLEAGTIFGVCVGEPMNQAVFKGIVPVITDEELMKDTPKVPFGYTKQMAEK 279
 DB 218 PPPQMAANKVNTVVGFCVGEPMNAQLVNOKIGYSALVGEIMKDHPEKAFSMQOWIRO 277
 QY 280 VNTYLAATKALIRAAIWLADNNKRNKEAIEMLAQOYVGADEVVLAASMNGTFEYE 339
 DB 278 NNNAQAAILMALLEAQOM--CDKAENKEBKMCISDRKYFNVAADIIIRAGKNDY--G 333
 QY 340 DGRALPDF---NTFFHGASYSYSAWVYLTQLRRWGMINEFKPDNYLDTKAKVYRPD 396

DB 334 DGRKQNFARHMKFADNMSYPKSHDIMEFLTEIDIMWGLPK---DTKQDVIYNQVNKED 390
 QY 397 IYLAAKELVAEKKAEPADTSIKPSQNFIDKVPEDANKENDYLAKPAI 449
 DB 391 LMKKAAR---AIGVADAE-IPASSS-RGVETFF-DGVKFDPEKREYELNSIKI 437

RESULT 8

AG1882
 nitrate transport ATP-binding protein nrtC [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C/Accession: AG1882
 R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AG1882
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-657 <KUR>
 A:Cross-references: UNIPROT:Q9Y276; GB:BA000019; PIDD:BA072568.1; PID:g17129956; GSPDB:G
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: nrtC
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 24.0%; Score 587; DB 2; Length 657;
 Best Local Similarity 35.8%; Pred. No. 4.6e-35;
 Matches 150; Conservative 74; Mismatches 139; Indels 56; Gaps 17;
 QY 39 KEDLKGFIKLTDMAPLVAAEKGFEEDEGL-FVQLEAQANKVVMDRVNGELDGSHTL 97
 DB 276 KAVIEIGFPLTDSAPLIVAEKGFPAKYGGLNVILNRAANNQALATAGVVTGLDAAGV 335
 QY 98 AAPPLAASVGGTADIEV--PFSMGFNCAITVSNIEIHQMKPNIPLEGKRVHPKAD 155
 DB 336 ACPMPALTITGASQPTPTPINALNLSRNNAITFSRLYNQGVRSI-----AD 383
 QY 156 YLKPVEKYKAEKGFPMNMTFPAGSHNITKLRYWLAAGINPGYSPPOISQIGADAL 215
 DB 384 -LKAVID--SSPDQILTLVAVSHASKNQILRYWLAAGID-----PDQVDS----- 427
 QY 216 LSVTPPQMPSTLEAGTIFGVCVGEPMNQAVFKGIVPVITDEELMKDTPKVPFGYTKQ 275
 DB 428 LTVIPTQVWSQLKGNIDGCAGEPMNYQAVHDDLGFAATALAEIWSQPKVIGVRD 487
 QY 276 VAEKIPNTYLAATKALIRAAIWLADNNKRNKEAIEMLAQOYVGADEVVLAASMNGTF 334
 DB 488 WAQKPEYILNIVKALIEACKY--CDLJNRREIIEILCRPEY---LDVNPAYVRSFPI 541
 QY 335 -EYKDD---KRALDQNFHFGASYSYSAWVYLTQLRRWGMINEFKPDNYLDTK 390
 DB 542 DRYDGDGTPPOQLTAIYNTLTKNTYPTILMTTQWMAWG-LTPF-PKOW--VEITTE 598
 QY 391 NVYRPDIYLAAKELVAEKKAEPADTSIKPSQNFIDKVPEDANKENDYLAKPAI 449
 DB 599 RVCRTDIFCAAARDL-----GLLDIGEDPI-----HLFDGLT-FNPSPIRYLAKSL 646

RESULT 9

AG1882
 NrtA-type periplasmic nitrate transport binding protein, probable Sma0585 [imported] - Sj
 C:Species: Sinorhizobium meliloti
 C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
 C/Accession: G95300
 R/Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowne, J.; Kaiman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.; Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A>Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
 A:Reference number: A95262; MUID:21396509; PMID:11481432

Best Local Similarity 32.7%; Pred. No. 1,4e-32;
Matches 148; Conservative 68; Mismatches 179; Indels 57; Gaps 16;

QY 6 RSSSKLLLTLSASLAW-GLTTPDVGAAGKLEKEDLKGFIKLTDMAPLAAAEKGF 64
DB 16 RSLMGAATLAAAKAFPGAHAAAGP-----EVAKARLGFTALDSSPLIATKERGLF 71
QY 65 EDEGL-FVQLEAQNKKVMDRV-----NGLDGSNMLAPAPLAAVGF---CTKADIV 116
DB 72 AKGMPIEVYKQASAPATRDNLVLAERGDIGAHLTPMPLMTGTITGAPTPMT 131
QY 117 PSMGFNGAIVTSNEIWHQMKENIPLEGKPVHPKADYLKEVVEKYAEKGFPMAMT 176
DB 132 LARLNTGGISIGNDL-KSVKGLNSAGA-----KAKEL-----QMKAGIAVAMT 179
QY 177 PFGASNIKLRWYLAAGINPGYSPPODISQIGADALLSTPPQPMSTLEAGTIFGY 236
DB 180 FPGTIDILIRWYLAAGINP-----DADVSTIIVPPQMANMKAQGTODAF 226
QY 237 CGEPPNQOAVFPGIGVPIITDELMKDPREKVGVTQMAEKVPMTYLAVTKALIRAAI 296
DB 227 CGEPPNQOAVNOQVITKCLTSELNNHPEKALGRASVDRYPRAAQITAAVQDAQ 286
QY 297 WLDADNNKREKAEIEMLAQKQYVADVEVLAASMGTFEYKDKRAL---PDFTFFPH 353
DB 287 W--CDKAAVLPGQCSIVSGKQYVNVPMGDLPRLOGTVDY--GDGRTLNKSPHRMKTAD 342
QY 354 GASYPSSAVALYIQLRWGMINEFKPDNWTIDTAKNYRPDIYLAAKEVAESKAKA 413
DB 343 NASFPKSHDLWFTLEDIRMGVLPQKTNKALVD--KVNRSDIWRAAKSIGSG---- 395
QY 414 EDFPADTSIKPSQNFIDKVPDPANKNDYLA 445
DB 396 ---PAGDS-RGVERFPDGKV-FDPANPGLYA 422

RESULT 12
558738
nitrate-binding protein nrtA precursor, periplasmic [similarity] - Phormidium lamosum
C:Species: Phormidium lamosum
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C:Accession: S58738; S56641; S62124
R:Merchan, F.; Friele, R.; Kindle, K.L.; Liama, M.J.; Serra, J.L.; Fernandez, E.
Plant Mol. Biol. 27, 1037-1042, 1995
A:Title: Isolation, sequence and expression in *Escherichia coli* of the nitrite reductase
A:Reference number: S56640; MUID:95284340; PMID:7766873
A:Accession: S56641
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-67, 'A', 69-442 <MEM>
A:Cross-references: EMBL:219598
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C:Genetics:
A:Gene: nrtA
C:Keywords: nitrate transport; periplasmic space

Query Match 22.3%; Score 545.5; DB 2; Length 442;
Best Local Similarity 33.3%; Pred. No. 2,9e-32;
Matches 145; Conservative 67; Mismatches 175; Indels 49; Gaps 14;

QY 27 IAP-DVGAVGKLEKEDLKGFIKLTDMAPLAAAEKGFEDGL-FVQLEAQNKKVMD 84
DB 44 LAPAPVAADTPEVITAKLGICPDGLADLIILAKEGLPFAKYGMPDVEVLKQASNAVTMD 103

QY 85 RVV-----NGLDGSNMLAPAPLAAVGFCTKADIVPFS-----MGFNAGIATVSNIRHQ 136
DB 104 NLESGGGGIGDAHILTPMPLMSAGITTKGQKVPKNIARLNVNGGICLANTY--- 160
QY 137 MKRPNIPLLEGKAPVHPKADYLKRVVEKYAEKGFPMAMTFFPAGSHNIKLRWYLAAGIN 196
DB 161 ----KDLKIGTSSPLKEAFKA-----KAEGERIAATVTFPGTIDLMRWLSSGID 211
QY 197 PGYSPPODISQIGADALLSTPPQPMSTLEAGTIFGYCGEPPNQOAVFPGIGVPI 256
DB 212 -----FDKDIS-----TIVPPQMANIKVNMMEFECGEPPADQVNGGLGYNAM 258
QY 257 TDELMKDPREKVGVTQMAEKVPMTYLAVTKALIRAAIWLDADNNKREKAEIEMLAQ 316
DB 259 TTGELMKDPEKAFARADWQHPKAAKALMAVQDAIW--CDPAPKEKQCVISGR 316
QY 317 QYVQADVEVLAASMGTFEYKDKRALPDFTFFHGASYSYSASVALYIQLRWGMNI 376
DB 317 EWFKPVVEDILERSKGNFLGVRLNSPLLMKFMNDASYPFKSHDLWFTLEDIRMGYL 376
QY 377 NEFKPDNWTIDTAKNYRPDIYLAAKEVAESKAKAEDFPADTSIKPSQNFIDKVPD 436
DB 377 PADDTKALVDA---VNREDLWREAKAIGQBA---IPASTS-RGVETFEF-DGVKPD 426
QY 437 AKRPNDYLAKEPAIGK 452
DB 427 PENPSAYLS--ALKIK 440

RESULT 13
AH2165
bicarbonate transport ATP-binding protein cmcC [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AH2165
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, S.; Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2165
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-667 <KUR>
A:Cross-references: UNIPROT:O8YT47; GB:BA000019; PIDN:BAB74578.1; PID:G17131973; GSPDB:GT
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: cmcC
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 22.0%; Score 538.5; DB 2; Length 667;
Best Local Similarity 33.3%; Pred. No. 1,7e-31;
Matches 140; Conservative 70; Mismatches 158; Indels 53; Gaps 14;

QY 30 DVGAVGKLEKEDLKGFIKLTDMAPLAAAEKGFEDGL-FVQLEAQNKKVMDRVN 88
DB 270 DVARRG-LEKWLLEIGFLPLTACAPLAAAEKGFETKGLDVNLVRETSMGIVDGKMG 328
QY 89 GELDGSNMLAPAPLAAVGFCTKADIV--PSMGFNAGIATVSNIRHQMKENIPLEGK 146
DB 328 GTYDAQNPSCMPMLWTJAGHDNQLPVYATLITMRNGAIIATKARFD-----EGV 380
QY 147 KPHPIKADYLKRVVEKYAEKGFPMAMTFFPAGSHNIKLRWYLAAGINPGYSPPODI 206
DB 381 RSLSPFK-NYL-----LRTDQRHIMGVHPASAHENILLRWLAAGIDF----- 424
QY 207 SCQIGADALLSTPPQPMSTLEAGTIFGYCGEPPNQOAVFKIGVPIITDELMKQTP 266
DB 425 ----DLADVMTIIPQAVADLQNKSIDGYCGEPPNRAAVENYGFITATDLEWLGHP 480
QY 267 EKVPFTQMAEKVPMTYLAVTKALIRAAIWLDADNNKREKAEIEMLAQKQYVADVEVL 326


```
Db 481 GATVLCREBMAERYPTHTALTKALLBACY--CSPRENVEVRRIYAGDVTDLDT 538
Qy 327 AASMNGTFEYKDKRALPDF--NTEFRHGA-SYPSYSAVWYLTOLRKGMINERKPDN 383
Db 539 QLEDPNSLVICDDI--HPLNDYAHHQFASASAINRPTBQIWMISQIARNG--DTEPPRN 594
Qy 384 WTLDTAKNVYRPDIYLAALKEVLAEGKAKAEDFPADTSIKPSONFFIDKPPDANKENDY 443
Db 595 W-VEEVERVCVRVVFSTAAREL-----GLDISYTRQPIELFDGTFFNADDPDIAY 642
Qy 444 L 444
Db 643 L 643

RESULT 14
S75959
nitrate transport protein C-1 - Synechocystis sp. (strain PCC 6803)
N/Alternate names: protein slr0043
C/Species: Synechocystis sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 16-Aug-2004
C/Accession: S75959
R/Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimp, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A/Reference number: S74322; MUID:97061201; PMID:8905231
A/Accession: S75959
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-667 <KAN>
A/Cross-references: UNIPROT:O55462; EMBL:D64006; GB:AB001339; NID:91001291; PIDN:BA01080
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C/Genetics:
A:Gene: nrcC-1
C:Superfamily: ATP-binding cassette homology
C:Keywords: ATP; nucleotide binding; P-loop
P/25-215/Domain: ATP-binding cassette homology <ABC>
P/42-49/Region: nucleotide-binding motif A (P-loop)

Query Match 21.9%; Score 534.5; DB 2; Length 667;
Best Local Similarity 31.7%; Pred. No. 3.3e-31;
Matches 146; Conservative 73; Mismatches 154; Indels 87; Gaps 16;

Qy 10 KULLTSLASLAWGLTTPDVGAVGKLEKEDLKFGFIKLTDMAPLVAAEKGFEDGL 69
Db 261 KLLRAKKTALIAHGH-----LKKVNLGLGVPLVACAPLVVAQEKGFPAKHGL 308
Qy 70 -FVLEAQAAMKVVMDRVVNGELDSHMLAPPLASVGFSTKADIEV--PFSMGFNGNA 126
Db 309 DEVSLVRETSWKGIVDGIAGVIGAGVAGQMPAGMPTWLTAGYREQSIPVVSALTMTXGNA 368
Qy 127 ITVSNIEIWMQKPNIPLEGKRPVPIKADYLKPVVEKYAEGKPFMMAMTFPAGSHNIX 186
Db 369 ITLSKLLDQ-----GIYTAE-DPRQLL--ASDGRHTLGVNHPSPENLILL 413
Qy 187 RTWLAAGGINPGYSPDIDISQIGDALLSVTPPOMSTLEAGTIFGVCVEEPNQA 246
Db 414 RYWLAAHNNP-----DRDVLKTIIPQMAVADLAKAGTIDGCVSEPPNLLA 460
Qy 247 VKGIGVPIVTDDELKQTPREKVGVTAKMAEKYPTVYLAVTALRAAIWLDADNNKOR 306
Db 461 SNEGAGFSLATDLEIQQNHGKVLGVREDAIAPHNTHALVALLEACAYC-ADPN-HE 518
Qy 307 KEAIEMLAQYVGADEVY-----AASMNGTFEYKDKRALPDFTEFRHGAAYP 358
Db 519 MEIRBELLATROYLSTIDYIHLGDPGRGTCRLGNPEY-----SHHLEFGDQFPNP 569
Qy 359 SYSSAVWYLTOLRKGMINERKPDNMYLTDTAKNVYRPDIYLAALKEV-----V 406
Db 570 SRTHELMNMTQARWGDIP--PFRNW-VEILERVCGVGFSTAARELGYDVNYQROPAL 626
```

```
Qy 407 AEGKAKAEDFPADTSIK-----PSQNFIDKVPFPANKP 440
Db 627 FDGKV-----FNADDPYAIYINQVTHRNFTAEVHLNPTP 662

RESULT 15
S77389
nitrate transport protein A-2 - Synechocystis sp. (strain PCC 6803)
N/Alternate names: nitrate transport 45K protein; protein slr1450
C/Species: Synechocystis sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C/Accession: S77389
R/Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimp, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A/Reference number: S74322; MUID:97061201; PMID:8905231
A/Accession: S77389
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-446 <KAN>
A/Cross-references: UNIPROT:P73452; EMBL:D90906; GB:AB001339; NID:91652492; PIDN:BA017492;
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C/Genetics:
A:Gene: nrcA-2

Query Match 21.5%; Score 525; DB 2; Length 446;
Best Local Similarity 28.5%; Pred. No. 9.3e-31;
Matches 139; Conservative 78; Mismatches 186; Indels 84; Gaps 15;

Qy 1 MKTIIRSSKKULLTSLASLAWGL-----TIAPVGAANGKL 37
Db 1 MSNFSRSTIRKKEFTYGA-AIGGVVLAGTSTTSTGTGSSSTDQALSPVBEENAP 59
Qy 38 EKEDLKFGFIKLTDMAPLVAAEKGFEDGL-FVLEAQAAMKVVMDRVV-----NGELD 92
Db 60 EYVTLALGFIALTDAPLIIAEKGFYAKGMPDEVLAQASGTTIRDNVLGASAGSID 119
Qy 93 GSHMLAPPLAASVGF--GTYKADIEVPPSMGFNGNAITVSNIEIWMQKPNIPLEGKPV 149
Db 120 GAHIIIPMLPLTMTGVTGDKPTPTYILARLNVGGQIQGANNY-----KDLKVGTD 172
Qy 150 HPYKADYLKPVVEKYAEGKPFMMAMTFPAGSHNIXRTWLAAGGINPGYSPDIDISQ 209
Db 173 APKAEFAVYTDK-----VAATFPQGTIDMIRYWLAAAGMEP-----211
Qy 210 IGADALLSVTPPOMSTLEAGTIFGVCVEEPNQAQVFGIGVPIVTDDELKQTPREKV 269
Db 212 -GMDPSTIVYPPAQMAANKVAMBSFCGEPPLDITVAGVGYQALITGQIKDHBEKA 270
Qy 270 FGVTQMAEKYPTVYLAVTALRAAIWLDADNNKKEAIEMLAQYVGADEVYLAAS 329
Db 271 FGRADWVQNPRAKALMAVDAQW--CDQAEKKEWCOQLTSKREMPKVPFEDIIDR 328
Qy 330 MNGTFEYK-----DKRALPDENTFFRGAASPSYSSAVWYLTOLRKGMINERKPDN 384
Db 329 SKGIVNFGNGQETFEQEIW--QKYWVNASYPYSHQWFLTENIRWGYLPAASDTTKA 385
Qy 385 YLDTAKNVYRPDIYLAALKEVLAEGKAKAEDFPADTSIKPSONF--FIDKVPDANKPN 442
Db 386 IYD--KVRREDIMRELAQAL-----EVRADQIPS-----SPSRKIEITFFDGIITDPENPQA 434
Qy 443 YLAKFAI 449
Db 435 YLDSLKI 441

Search completed: May 18, 2005, 17:43:45
Job time : 60 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2005, 17:36:47 ; Search time 179 Seconds
(without alignments)
1327.401 Million cell updates/sec

Title: US-10-689-200-2

Perfect score: 2443
Sequence: 1 MKTIRSSSKLLTLTSLASL.....AKFALGKQGVAGKVVVD 464

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: uniprot_spprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1597	65.4	462	Q8KY55	azospirillum
2	1586	64.9	475	Q7MDT5	7mdt5 vibrio vuln
3	1572	64.3	487	Q8D6X8	08d6x8 vibrio vuln
4	1564	64.0	454	Q87HA7	087ha7 vibrio para
5	1257.5	51.5	553	Q7UYV1	07uyv1 rhodospirillum
6	651	26.6	670	NRTC_SYNY3	73450 synecocyst
7	627.5	25.7	430	Q8U925	08u925 agrobacteri
8	616.5	25.2	441	Q92VK8	092vk8 rhizobium m
9	611.5	25.0	431	Q8U924	08u924 agrobacteri
10	610.5	25.0	440	NRTA_ANASP	044252 anabaena sp
11	610.5	25.0	440	Q7NEC4	07nec4 gloeobacter
12	610	25.0	429	Q7NEC4	07nec4 gloeobacter
13	594	24.3	427	Q7NKB1	07nkb1 anabaena sp
14	587	24.0	457	Q8Y2F6	08y2f6 ralbactonia s
15	585.5	24.0	420	Q8Y2F5	08y2f5 ralbactonia s
16	582.5	23.8	439	Q898F9	089f9 bradyrhizob
17	578.5	23.7	412	Q898H4	089h4 rhizobium 1
18	578	23.7	420	Q6D2V0	06d2v0 erwina car
19	577	23.6	439	Q8DJ78	08dj78 synecococc
20	576	23.6	430	Q922Z1	0922z1 rhizobium m
21	575.5	23.6	418	Q48466	048466 klebsiella
22	572	23.4	443	NRTA_SYNP7	038063 synecococc
23	567	23.2	454	Q8DJ76	08dj76 synecococc
24	566	23.2	425	Q98H12	098h12 rhizobium 1
25	559.5	22.9	388	Q89JH2	089jh2 bradyrhizob
26	549.5	22.5	430	Q9AA18	09aa18 caulobacter
27	548.5	22.5	448	Q7NUI6	07niu6 gloeobacter
28	548.5	22.5	458	Q8DH7	08dh7 synecococc
29	548	22.4	428	Q6WRT2	06wrt2 rhodobacter
30	545.5	22.3	442	Q51880	051880 photomidium
31	542.5	22.2	666	Q7NUI3	07niu3 gloeobacter

32	540	22.1	655	Q7NKA9	07nka9 gloeobacter
33	538.5	22.0	667	Q8YT47	08yt47 anabaena sp
34	536.5	22.0	653	Q55107	055107 synecococc
35	534.5	21.9	667	Q55462	055462 synecocyst
36	531	21.7	404	Q88141	088141 pseudomonas
37	525.5	21.5	462	Q6N722	06n722 rhodopseud
38	525	21.5	403	Q883P3	0883p3 pseudomonas
39	525	21.5	446	NRTA_SYNY3	073452 synecocyst
40	521.5	21.3	533	Q6WIF8	06wif8 synecococc
41	510.5	20.9	459	Q89JF5	089jf5 bradyrhizob
42	509.5	20.9	559	Q7U3E5	07u3e5 synecococc
43	506	20.7	386	Q6N3I6	06n3i6 rhodopseud
44	502.5	20.6	458	Q8IT49	08it49 anabaena sp
45	494	20.2	402	Q912V6	0912v6 pseudomonas

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	462 AA.
ID Q8KY55			
AC Q8KY55;			
DT 01-OCT-2002 (TREMBlrel. 22, Created)			
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)			
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)			
GN Nitrate transporter substrate-binding protein.			
OS Acetivibrio brasiliense.			
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;			
OC Rhodospirillaceae; Azospirillum.			
OX NCBI TaxID=192;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=9245;			
RA Steenhouck O., Ptacek D., Verreth C., Vanderleyden J.;			
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AF283498; AAM73543.1; -			
DR InterPro; IPR002110; ANK.			
DR PRINTS; PR01415; ANKYRIN.			
SQ SEQUENCE 462 AA; 50270 MW; 25C86DF070421EBC CRC64;			
Query Match	65.4%;	Score 1597;	DB 2; Length 462;
Best Local Similarity	65.6%;	Pred. No. 6.3e-112;	
Matches 301;	Conservative 60;	Mismatches 92;	Indels 6; Gaps 4;
QY	7	SSSKLLTLTSLASLAWGLTIAPDVAGVGLKEDLKRGIKLTMDAPLAVALAEKFFED 66	
DB	6	ASPRFALLSMAATLA--LMDGSAQAAPLDVEKDQKLGFTKLTMDAPLAIAEKGFED 62	
QY	67	BGLFVQLEAQAAMKVVMDRVNVELDGSMLAPAPLAASVGFQTKADIEVPFSGFGNA 126	
DB	63	BGLSVTLLEPQANKKVLIDRYISGELDGAHMLAQPIGATITGFTQANVTAFSMDLNGK 122	
QY	127	ITVSNIEIWMQKKNIPPL-BGGKPVHPIKADYLKPVVEKYKAEKPPNMAWTFPAGSHIK 185	
DB	123	ITLSNVEWEMKKNLKPDPGKPIHPIKADALKPVIAQRAEKKPFTMGVFPVSTHNYE 182	
QY	186	LRYWLAAGGINSPIYSPPODISQIAGDALLSTPPPPQNPSTIEAGTTFICYCGEPNQQ 245	
DB	183	LRYWLAAGGINSPIYSA-PDVDSQIQADALLSTPPPPQNPSTIEAGTTFICYSGEPNQQ 241	
QY	246	AVFGKIGVPIITDEELMKDTPPEKVPFVTQMAKRYNTYLAATKALIRAIWIDADNNK 305	
DB	242	AVFGKIGVPIITDEELMKDTPPEKVPFVTQMAKRYNTYLAATKALIRAIWIDADNNK 301	
QY	306	RKEAIEMLAQOQYVADVEVLAASNNGTPEYKEDKRALPDEFNTPFRHGAASYPSYSAW 365	
DB	302	RAEAVIKLAKSEVYVADAVIANSMGTFEYKEDGRAVDDFVFRVYATYFYSDAW 361	
QY	366	YLTQLRKGKINLFFKPDNWTLDPAKVYRPDIYLAALAKSLVAGKKAEDFP-ADTSIIP 424	
DB	362	YLTQMRWQIEMAKPDADYDEYARVYRREIYLAARLLVEGKAKADFPWTSIDGYXP 421	

QY 425 SONEFIDKVPFDANKPNDYLAKEAIGLKQYVAGKGV 463
 DB 422 LDNGFIDGLAYDGRKRENYLTKLPIGLKGQAVQGGQLV 460

RESULT 2

Q7MDT5 PRELIMINARY; PRT; 475 AA.
 AC Q7MDT5;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Nitrate transporter system, periplasmic component.
 GN OrderedLocustNames=VVA0951;
 OS *Vibrio vulnificus* (strain Y016).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; *Vibrio*.
 OX NCBI_TaxId=196600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14656965; DOI=10.1101/gr.1295503;
 RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
 RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
 RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
 RT "Comparative genome analysis of *Vibrio vulnificus*, a marine
 pathogen.";
 RL Genome Res. 13:2577-2587(2003).
 DR EMBL; AP005348; BAC96977.1; -.
 KW Complete proteome.
 SQ SEQUENCE 475 AA; 53801 MW; B61E6D1915DC3C67 CRC64;

Query Match 64.3%; Score 1586; DB 2; Length 475;
 Best Local Similarity 65.6%; Pred. No. 4.4e-111;

Matches 292; Conservative 60; Mismatches 85; Indels 8; Gaps 5;

QY 20 LVMGTLTPDV--GAVGKLEKEDLKRFKLTDMAPLVAEKGFEDEGLFYQLEAQA 77
 DB 32 LLSW-LAFRPPVLAEGEAEKEDLRFKLTDMAPLVAEKGFEDEGLFYVLEAQA 90
 QY 78 NKKVMDRVVNGELDSHMLAPAPLAASVGFGRKADIEVPFSMGFNATITVSNETHOM 137
 DB 91 NKKVLLDRVIDGLDGAHMLAGPPLGATIGIGTQAVITAFSMDLNGNATITVSNVWQOM 150
 QY 138 KENIPLEG-GKPVHPKADYLKPVVEKYAKGKPFMMATFPAGSHNIKRLWLAAGIN 196
 DB 151 KENLAQSDGKPVHPKADALKPVVTSYRDQKAFNMGWFPVSTHNYELRWLAAGIH 210
 QY 197 PGYSPPO-DISGOIGADALLSTPPQMPSTLEAGTIFGTCVGEPMNQOAVRKGIVPV 255
 DB 211 PGYTAHKGDSQINADVLLSTPPQMPATWEAGTIGTCVGEPMNQOAVRKGIVPV 270
 QY 256 ITDEELMKDPEKVFQVTKQMAEKYPTNYLAATKALIRAAIWLADNNKRRKEAIBMLAQ 315
 DB 271 VTDYEITKKNPEKVFQVADMAEKYPTNTHIRVYKALIRAAHMLDENNNRTEAIVCLSR 330
 QY 316 KQYGVADVEVLAASNGTFFYEKDKRALPDENTFFRHGASYSYSAAVYLTQLRRWGM 375
 DB 331 SEYVGADAEVLAANSMTGTFFYEKDKRQVPDEFVFRYNAITYYSDAIWLQLQRRWGM 390
 QY 376 INEFKPDNMYLDTAKVYRPDIYLAAKELVAGKAKADFP--ADTSIKSQNFIDK 432
 DB 391 IEHQKPDNSWYMDIAKQVYRPDIYQRAAEALIEBGLTASDFDFAKESGFRPQTHFIDQ 450
 QY 433 VPEIDANKPNDYLAKEAIGLKQYV 457
 DB 451 IHYDGRSPNAYLQGFALIGKGSSEL 475

RESULT 3
 Q8D6X8 PRELIMINARY; PRT; 487 AA.
 AC Q8D6X8;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE ABC-type nitrate/sulfonate/bicarbonate transport system, periplasmic component.
 GN OrderedLocustNames=VW20393;
 OS *Vibrio vulnificus*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; *Vibrio*.
 OX NCBI_TaxId=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.B.;
 RT "Complete genome sequence of *Vibrio vulnificus* CMCP6.";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016809; AA007349.1; -.
 KW Complete proteome.
 SQ SEQUENCE 487 AA; 54515 MW; PESD5F11C5B865F4 CRC64;

Query Match 64.3%; Score 1572; DB 2; Length 487;
 Best Local Similarity 65.1%; Pred. No. 5.2e-110;
 Matches 289; Conservative 63; Mismatches 86; Indels 6; Gaps 4;

QY 20 LVMGTLTPDVGA-VGKLEKEDLKRFKLTDMAPLVAEKGFEDEGLFYQLEAQA 78
 DB 44 LLSWLAFAFPVLAEGEAEKEDLKRFKLTDMAPLVAEKGFEDEGLFYVLEAQA 103
 QY 79 WKVMDRVVNGELDSHMLAPAPLAASVGFGRKADIEVPFSMGFNATITVSNETHOM 138
 DB 104 WKVLLDRVIDGLDGAHMLAGPPLGATIGIGTQAVITAFSMDLNGNATITVSNVWQOM 163
 QY 139 PNIPLEG-GKPVHPKADYLKPVVEKYAKGKPFMMATFPAGSHNIKRLWLAAGINP 197
 DB 164 PNLVQSDGKPVHPKADALKPVVTSYRDQKAFNMGWFPVSTHNYELRWLAAGIH 223
 QY 198 GYSPPO-DISGOIGADALLSTPPQMPSTLEAGTIFGTCVGEPMNQOAVRKGIVPV 256
 DB 224 GYVAPHKGDSQINADVLLSTPPQMPATWEAGTIGTCVGEPMNQOAVRKGIVPV 283
 QY 257 TDEELMKDPEKVFQVTKQMAEKYPTNYLAATKALIRAAIWLADNNKRRKEAIBMLAQ 316
 DB 284 TDEYETKKNPEKVFQVADMAEKYPTNTHIRVYKALIRAAHMLDENNNRTEAIVCLSR 343
 QY 317 QYGVADVEVLAASNGTFFYEKDKRALPDENTFFRHGASYSYSAAVYLTQLRRWGM 376
 DB 344 EYVGADAEVLAANSMTGTFFYEKDKRQVPDEFVFRYNAITYYSDAIWLQLQRRWGM 403
 QY 377 NEFKPDNMYLDTAKVYRPDIYLAAKELVAGKAKADFP--ADTSIKSQNFIDK 433
 DB 404 EHOKPDNSWYMDIAKQVYRPDIYQRAAEALIEBGLTASDFDFAKESGFRPQTHFIDQ 463
 QY 434 PEDANKPNDYLAKEAIGLKQYV 457
 DB 464 RYDGRSPNAYLQGFALIGKGSSEL 487

RESULT 4

Q87HA7 PRELIMINARY; PRT; 454 AA.
 AC Q87HA7;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative nitrate transport protein.
 GN OrderedLocustNames=VPA1058;
 OS *Vibrio parahaemolyticus*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; *Vibrio*.
 OX NCBI_TaxId=670;
 RN [1]
 RP SEQUENCE FROM N.A.

DR GO: 00005524; F:ATP binding; IEA.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 553 AA; 60761 MW; 743DPAC3267B5D63 CRC64;

Query Match 51.5%; Score 1257.5; DB 2; Length 553;
Best Local Similarity 50.7%; Pred. No. 3,28-86;
Matches 249; Conservative 74; Mismatches 133; Indels 35; Gaps 8;

QY 2 KTIIRSSSKLLLTLSASLAW-----GLTIADVGAVGK----- 36
DB 62 KKAIRTNKMLLGGCVALLAAVFTGCADSGVSLIEDLEAAAKVDISKIEVDVNPTESAAML 121
QY 37 -LEKEDLKFGKIKITDAPFLAVAAEKFFPEBEGFLVLEQAQAMKVMYMDRVNSELDSH 95
DB 122 DLEKSDLTFFGRIKLTDAPLVIAEKKGYPFDEGLVLTLEQSNMKILLDVIYNQLDGAH 181
QY 96 MLAPAPLAASVGFCTKADIEVFPMSGNGNAITVSNELTHOMKNIP-LEGGKRVHPKA 154
DB 182 MLAQPIGATIGVGSGSPIYATVSLDYNNGCITVSNEMWQMOMQENDELKSPYTKHFISA 241
QY 155 DYLRPVVEKKYAA-GKPPNNAMTFPAGSHNIKRMYLAAGINPGYSSPPDISGQICAD 213
DB 242 ASLRPIYEDVLDQDGEFFPGMGPVSTHNEYRMYLAASGHGQWT-ESDINGPFDQAQ 300
QY 214 ALLSVTPPPQPMSTLEKGTIFGCVGEPNMQAVFKIGVPVITDELMKDTPEKVFQVT 273
DB 301 VKLSVTPPPQPMQLEADIVGVGCVGSPNMQKAVVTSIGVPTVNTYIMKNPEKVFQVT 360
QY 274 KQMAEKYPNMYLATATKALISAATMLDA--DNMK--NKEKVEMLAQOYVQADVETLAAS 329
DB 361 GEMAEKMPQTHLAIVKALIRAKWLDATDQSGKLVNREAEVETLSRDYVGAEREVIGNS 420
QY 330 MNGFEVEKDKRALPDNTEFFRHGAGVPSYSASVWLTQLRWGMINEFPKDMWYLDTA 389
DB 421 MNGFVFGQSDYRMPDPNVPFFKHABAYPHYSDAIWLTLQMRKQGQTESGASVAYETA 480
QY 390 KNVYRPDIYLAAEKELVAEGKAKADEPAD--TSIKSQNFIFDKVDPDANKPNDYLAKE 447
DB 481 KKIYKPEIRYQAABELLISEGKLDPNEIIPADPDGYRAVTFEFDGNKYDADPDIGYINSF 540
QY 448 AIGAKGQYTA 458
DB 541 EIGNKDDESLA 551

RESULT 6
NRTC_SYNY3 STANDARD; PRT; 670 AA.
AC P73450;
ID NRTC_SYNY3
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Nitrate transport ATP-binding protein nrtc.
OS Synecocystis sp. (strain PCC 6803)
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystie.
NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugisura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -I- FUNCTION: Probably part of a high-affinity binding-protein-
CC dependent transport system for nitrate. Probably responsible for
CC energy coupling to the transport system.
CC -I- SUBCELLULAR LOCATION: Membrane-associated (Potential).
CC

CC -1- SIMILARITY: Belongs to the ABC transporter family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
CC or send an email to license@1sb-sib.ch).
DR EMBL: D90906; BAA17490.1; -.
DR PIR: S77387; S77387.
DR HSSP: O58206; 1L2T.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transporter.
DR InterPro: IPR005890; NtrCD.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfam: TIGR01184; ntrCD; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome; Membrane; Nitrate assimilation;
KW Transport.
KM DOMAIN 1 254 ABC transporter.
FT DOMAIN 255 278 Linker.
FT DOMAIN 279 670 NtrA-like.
FT NP_BIND 42 49 ATP (potential).
SQ SEQUENCE 670 AA; 75100 MW; 03B47B6C7918AD14 CRC64;

Query Match 26.6%; Score 651; DB 1; Length 670;
Best Local Similarity 36.3%; Pred. No. 2.3e-40;
Matches 158; Conservative 68; Mismatches 153; Indels 56; Gaps 13;

QY 28 ADVGAVGK-----LEKEDLKFGFIKLTDMAPLAVAEKGFPEDEGL-FVQLEAQA 77
DB 260 AKKVGAVSQFAEAMGNGLEKINDLGFILPDCAFLVAKKEGFQKGLBEVNLVKEP 319
QY 78 NKKVVMDRVNGELGDSHMLAPAPLAASVGFTRADIEV--PSSMGFNAGLITVSEIWH 135
DB 320 SMOALDGI RERLDGACQVAGWPLALTITGMGSKTPLPMTAVMWSRNGNAILTSKKFAB 379
QY 136 QKKPNPLBEGKRVHPIKADYLKPVVEKTKABGKPPNMAFPPAGSHNIKLRWYLAAGI 195
DB 380 -----AGVKTLEDLR-----LKLAEPPDQVSTLGVHPASNONLLIRWYLAAGSI 424
QY 196 NPGVYSPPODISGQIGADALLSVTPPOMSTLEAGTIFGYCVGEPMNOQAVFKGIGVPE 255
DB 425 D-----PDQDIN-----LWLPPEPQVNSNLEAGNIDGFCVGEPMNSVYAKQNLGYVI 471
QY 256 ITDEELMKDTPPEKVFQVTQMAEKYPNTYLAATKALIRAAIWLADNNKREALEIMLAQ 315
DB 472 ATCDLIDWNGHPEKVLGMBREBWKYPATHLALVALLACEY--CDRRHRRQGIIDLYAL 529
QY 316 KOYVGADVEVLAASNGTFEYEDK-DKRALPDENTFRRGASVPSYSSAVWYLTQLRNG 374
DB 530 PDIYGVSTIYISPGFTLEYDQGNDAEMLIDFNQYVKSQNSPSSBSGLMILTQLRWG 589
QY 375 MINEFKPDWYLDATKNVPRDIYLAARELVAEKAKAEDPADTSIKPSQNFIDKVP 434
DB 590 YID--FPKRW-VBIIRVRRPDLFGAACHL--GMPDLEGDHNVSL-----FDGAV 636
QY 435 FDANKENDYLAAPAI 449
DB 637 FTPNDPLGYIKRFTI 651

RESULT 7
Q8U925 PRELIMINARY; PRT; 430 AA.
AC Q8U925;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE ABC transporter, substrate binding protein.
GN Name=ntrLA; OrderedLocustNames=Atc3905;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium; Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dupont;
RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;
RA Wood D.W., Serubal J.C., Kaul R., Monks D.E., Ktajiama J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Cleedenning J., Deatherage G., Gilliet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmer A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323 (2001).
DR EMBL: AB009321; AAL44713.1; -.
DR PIR: AC03037; AC03037.
DR PIR: H98248; H98248.
DR TIGRfam: TIGR01409; TAT_signal_seq; 1.
KW Complete proteome.
SQ SEQUENCE 430 AA; 46766 MW; 0388162B02B2D8C CRC64;

Query Match 25.7%; Score 627.5; DB 2; Length 430;
Best Local Similarity 35.5%; Pred. No. 7.7e-39;
Matches 166; Conservative 70; Mismatches 171; Indels 61; Gaps 19;

QY 1 MTTIIRSS-SKDLU--TISASLAVWGLTIAPD--VGAQKLEKEDLKFGFIKLTDMAPL 55
DB 1 MKKIFSGIVSRRTILKTTATYALVTAVRTAPSGAPAAAEVGVAKAGIPLALTDAFL 60
QY 56 AVAAEKGFPEDEGL-FVQLEAQAANKVMDRVVNG-----ELDGSMLAPAPLAASVGF 110
DB 61 IIAEKGLFAKKGMPBEVELKQASWCATRDNLVLGASGIDGAIHLTPMYLMTGKYT 120
QY 111 KADIEVPS--WGFNGNALTIVSEIWHQMKPNPLBEGKRVHPIKADYLKPVVEKYA 166
DB 121 QNNVPEPMTILARLINDSGISVAKK--YAETGVOLDASK-----LKAAFEKKKA 168
QY 167 EKGPPNMAFPPAGSHNIKLRWYLAAGINPGVSPPODISGQIGADALLSVTPPOMS 226
DB 169 DGEIKAMATFPGGTHDMLIRYLAAGID-----PDQVDS-----TIVPPQOMA 215
QY 227 TLEAGTIFGYCVGEPMNOQAVFKGIGVPIYTDLELMKDTPEKVFQVTQMAEKYPNTYLA 286
DB 216 NKKVGMDFCVGEPMNEBOLVNOGIGFTACTTGELMKHPEALGMBADWVENKPNATGA 275
QY 287 VTKALIRAAIWLADNNKREALEIMLAQKQYVGADVEVLAASNGTFEYEDKRALPD 346
DB 276 LIMAWEAQW--CEBMAKEMSTILGRQFNVPKQVGLRLKGNINY--GNGRVLE- 330
QY 347 FNT-----FFRGASVPSYSSAVWYLTQLRNGMNEFEEDWYLDATKNVPRDIYLA 401
DB 331 -NTGLQMKFWQDHAAYPFPHSHDSWFTTENIRWG--KFAPDIDYKALVAKVREDIWRRA 386
QY 402 AKELVAEKAKAEDPADTSIKPSQNFIDKVPFDANKENDYLAAPAI 449
DB 387 AKDL---GVA---DLPASTS-RGKETFPDGKY-FDPENPSATLBSLSI 426

RESULT 8
Q7CTL1 PRELIMINARY; PRT; 471 AA.
AC Q7CTL1;
DT 05-JUL-2004 (TREMBlrel. 27, Created)

DT 05-JUL-2004 (TEMBLrel. 27, last sequence update)
 DT 05-JUL-2004 (TEMBLrel. 27, last annotation update)
 DE AGR_L_1886p.
 GN OrderedLocuNames=AGR_L_1886;
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 NCBI_TaxID=176299;
 RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
 RC STRAIN=Cereon;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Gutillo B., Goldman B.S., Cao Y., Askenazi M., Hailing C., Mullin L.,
 RA Houtrel K., Gordon J., Vaudin M., Iarchouk O., Bp A., Liu F.,
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappae C., Markelz B.,
 RA Flanagan C., Crowell C., Gerson J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.,
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58."
 RL Science 294:2323-2328(2001).
 DR EMBL; AEO08293; AAK89514.1; -
 DR InterPro; IPR006073; HSC20_C.
 DR TrIRPMS; TIGR01409; TAT_signal_seg; 1.
 DR TIGRPM; TIGR01409; TAT_signal_seg; 1.
 SQ SEQUENCE 471 AA; 51726 MW; 2650A5968C39D79A CRC64;
 Query Match 25.7%; Score 627.5; DB 2; Length 471;
 Best Local Similarity 35.5%; Pred. No. 8.7e-39;
 Matches 166; Conservative 70; Mismatches 171; Indels 61; Gaps 19;
 QY 1 MKTIIRSS-SKLLLT--TSSASLAWGLTIAPD--VGAVGLKEKEDLKFGFIKLTMAPL 55
 DB 42 MKKIFSGTYSRRITLKTATATVATPAFPGAFATAPREVKAQKGFIALTDAPL 101
 QY 56 AVAAEGFPEDEGL-FVQLEAQNKKVMDRVYNG---ELDSHMLAPPLAASVGFCT 110
 DB 102 IIAAEKGLPAKGMPEVEYLKQASMGATRDNLVGGASNGIDGAIHLITPMPYLMHGTXT 161
 QY 111 KADIEVPFS---MGFNMAITVSNELWQMKNPILGEGKPHPIKADVLKVEKYKA 166
 DB 162 QNNVPVPMITLARIINDSGISVAKK---YAEVGQVLDASK-----LKAPEKKKA 209
 QY 167 EGKPFPMAMTFPAGSHNKLRYLWLAAGINPGYSPPODISQIGADALLVTPPQMP 226
 DB 210 DGEIQAANTFPGCTHDWIRYLAAGID----FDKQVS-----TIIVPPQMA 256
 QY 227 TLEAGTIFGYCGEPPNQAQVFKGIVPTDELMKQTPKVFQVTKQMAEKYPTTYLA 286
 DB 257 NKKVGMNDFCVGEPPNNEQLVNGIGFTACTTGELMKGHEKALGMRADVVEKPNATKA 316
 QY 287 VTKALIRALWLDADNNKRNKEALEMLAQKQYGVADVEYLAASNGTFEYKDKALPD 346
 DB 317 LIMAWEAQQM--CDMANKEEMSTILGKQFNVPPKQVLRKLNIN--GNGRVLE- 371
 QY 347 FNT-----FFRHGASYSSAVVYLTLRRQWMEFKPDNMYLDTAKKVVPPDIYLA 401
 DB 372 -NTGLQMKWQDHASYPFHSHDSMTITENIRKQ---KFAPTDYKALVAKVREDIWRRA 427
 QY 402 AELVAEGKAKADEFPADTSIKPSQNFIDKVPDANKPNDYLAAPAI 449
 DB 428 AKDL---GVA---DLPASTS-RKETFPDQKV-FDPENPASYLESISI 467
 RESULT 9
 QY2VK8 PRELIMINARY; PRT; 441 AA.
 AC Q92VK8;
 DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-DEC-2001 (TEMBLrel. 19, last sequence update)
 DT 01-JUN-2003 (TEMBLrel. 24, last annotation update)
 DE Putative nitrate transport protein.
 GN ORFNames=SMD21114;

OS Rhizobium meliotti (Sinorhizobium meliotti).
 OC Plasmid pSymb.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 NCBI_TaxID=382;
 RX MEDLINE=21396508; PubMed=11481431; DOI=10.1073/pnas.161294698;
 RC STRAIN=1021;
 RA Finan T.M., Weidner S., Wong K., Buhmester J., Chain P.,
 RA Vorholzer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouy J.,
 RA Golding B., Puhler A.,
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliotti."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 DR EMBL; AL591985; CAC49093.1; -
 DR PIR; E95928; B95928.
 KM Complete proteome; Plasmid.
 SQ SEQUENCE 441 AA; 47208 MW; 2F2F4498B8F81D58 CRC64;
 Query Match 25.2%; Score 616.5; DB 2; Length 441;
 Best Local Similarity 33.8%; Pred. No. 5.4e-38;
 Matches 153; Conservative 74; Mismatches 182; Indels 43; Gaps 12;
 QY 19 SLAWGLTIAPDVGAVGLKEKEDLKFGFIKLTMAPLVAEKGFEDEGLFVQLEAQN 78
 DB 19 NLSAGALPAP--GSIRHSGSRTRVAGFIPVDAVLAIAAEFGAQRGEGITELVQDVS 76
 QY 79 WKVMDRVNNGELDSHMLAPPLAASVGFCTADIEV--PESMGFNMAITVSNELWQ 137
 DB 77 WANVRDLAFRQPDVAHMLSPMVAAMLGSGNSPSTIAPFSGRGGNATLSTRYGLM 136
 QY 138 KPNIPLEGKVPVPIKADYLKPYVEKYKAEK--PFNMAMTFPAGSHNKLRYLWLAAGIN 196
 DB 137 QQAGGLGCGEDALK-NAKXLAIVIRSAAGAARPLTGLTYPPSSINVERRYLAAGID 195
 QY 197 PGYSPPQDISQIGADALLVTPPQMPSTLEAGTIFGYCGEPPNQAQVFKGIVPT 256
 DB 196 P-----DSDVTLVVPVPMITLARIINDSGISVAKK---YAEVGQVLDASK-----LKAPEKKKA 242
 QY 257 TDELMKQTPKVFQVTKQMAEKYPTTYLAATVTKALIRALWLDADNNKRNKEALEMLAQK 316
 DB 243 TKQDIPSAPEKYLGNRPMAEARNRTVSRVLVVALDRAARW--SDPAPNAGRLAEVLAER 300
 QY 317 QYGVADVEYLAASNGTFEYKED-DKRALPDENVTFRRHGASYSSAVVYLTLRRQW 375
 DB 301 RHYAAPADIIIRYLAAGEFAIDEGNRRVBNYVFAAGRANYPBQALMTYSQMRMQ 360
 QY 376 INEFPDNNYLDTAKKVVPPDIYLAAKELVAEGKAKADEFPADTSI-----KPSQNF 430
 DB 361 TSLRQ---RVDAISAYRPDLVREAL-----GK---DALPADADAGRLEGATDGRFM 408
 QY 431 DKVPDANKPNDYLAAPAIGLKQTVVAGKV 462
 DB 409 DGHVPDGRIDYIKSF-----GAPSVNRSV 435
 RESULT 10
 QY2VK8 PRELIMINARY; PRT; 431 AA.
 AC Q8U924;
 DT 01-JUN-2002 (TEMBLrel. 21, Created)
 DT 01-JUN-2002 (TEMBLrel. 21, last sequence update)
 DT 25-OCT-2004 (TEMBLrel. 28, last annotation update)
 DE ABC transporter, nucleotide binding/ATPase protein
 DE (AGR_L_188191p).
 GN Name=ntC; OrderedLocuNames=AGR_L_188191; Atu3906;
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 NCBI_TaxID=176299;
 RX MEDLINE=21396508; PubMed=11481431; DOI=10.1073/pnas.161294698;
 RC STRAIN=1021;
 RA Finan T.M., Weidner S., Wong K., Buhmester J., Chain P.,
 RA Vorholzer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouy J.,
 RA Golding B., Puhler A.,
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliotti."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 DR EMBL; AL591985; CAC49093.1; -
 DR PIR; E95928; B95928.
 KM Complete proteome; Plasmid.
 SQ SEQUENCE 441 AA; 47208 MW; 2F2F4498B8F81D58 CRC64;
 Query Match 25.2%; Score 616.5; DB 2; Length 441;
 Best Local Similarity 33.8%; Pred. No. 5.4e-38;
 Matches 153; Conservative 74; Mismatches 182; Indels 43; Gaps 12;
 QY 19 SLAWGLTIAPDVGAVGLKEKEDLKFGFIKLTMAPLVAEKGFEDEGLFVQLEAQN 78
 DB 19 NLSAGALPAP--GSIRHSGSRTRVAGFIPVDAVLAIAAEFGAQRGEGITELVQDVS 76
 QY 79 WKVMDRVNNGELDSHMLAPPLAASVGFCTADIEV--PESMGFNMAITVSNELWQ 137
 DB 77 WANVRDLAFRQPDVAHMLSPMVAAMLGSGNSPSTIAPFSGRGGNATLSTRYGLM 136
 QY 138 KPNIPLEGKVPVPIKADYLKPYVEKYKAEK--PFNMAMTFPAGSHNKLRYLWLAAGIN 196
 DB 137 QQAGGLGCGEDALK-NAKXLAIVIRSAAGAARPLTGLTYPPSSINVERRYLAAGID 195
 QY 197 PGYSPPQDISQIGADALLVTPPQMPSTLEAGTIFGYCGEPPNQAQVFKGIVPT 256
 DB 196 P-----DSDVTLVVPVPMITLARIINDSGISVAKK---YAEVGQVLDASK-----LKAPEKKKA 242
 QY 257 TDELMKQTPKVFQVTKQMAEKYPTTYLAATVTKALIRALWLDADNNKRNKEALEMLAQK 316
 DB 243 TKQDIPSAPEKYLGNRPMAEARNRTVSRVLVVALDRAARW--SDPAPNAGRLAEVLAER 300
 QY 317 QYGVADVEYLAASNGTFEYKED-DKRALPDENVTFRRHGASYSSAVVYLTLRRQW 375
 DB 301 RHYAAPADIIIRYLAAGEFAIDEGNRRVBNYVFAAGRANYPBQALMTYSQMRMQ 360
 QY 376 INEFPDNNYLDTAKKVVPPDIYLAAKELVAEGKAKADEFPADTSI-----KPSQNF 430
 DB 361 TSLRQ---RVDAISAYRPDLVREAL-----GK---DALPADADAGRLEGATDGRFM 408
 QY 431 DKVPDANKPNDYLAAPAIGLKQTVVAGKV 462
 DB 409 DGHVPDGRIDYIKSF-----GAPSVNRSV 435
 RESULT 10
 QY2VK8 PRELIMINARY; PRT; 431 AA.
 AC Q8U924;
 DT 01-JUN-2002 (TEMBLrel. 21, Created)
 DT 01-JUN-2002 (TEMBLrel. 21, last sequence update)
 DT 25-OCT-2004 (TEMBLrel. 28, last annotation update)
 DE ABC transporter, nucleotide binding/ATPase protein
 DE (AGR_L_188191p).
 GN Name=ntC; OrderedLocuNames=AGR_L_188191; Atu3906;
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 NCBI_TaxID=176299;
 RX MEDLINE=21396508; PubMed=11481431; DOI=10.1073/pnas.161294698;
 RC STRAIN=1021;
 RA Finan T.M., Weidner S., Wong K., Buhmester J., Chain P.,
 RA Vorholzer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouy J.,
 RA Golding B., Puhler A.,
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliotti."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 DR EMBL; AL591985; CAC49093.1; -
 DR PIR; E95928; B95928.
 KM Complete proteome; Plasmid.
 SQ SEQUENCE 441 AA; 47208 MW; 2F2F4498B8F81D58 CRC64;

```

RC STRAIN=DuPont; PubMed=11743193; DOI=10.1126/science.1066804;
RX MEDLINE=21608550;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Giller W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland G., Palmeri A.,
RA Raymond C., Rouse G., Saenphumachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Ioo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58 ";
RL Science 294:2317-2323 (2001).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Gurolo B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Hounmel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Dougherty D., Scott C., Lapas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL; AEO09321; AAL4714.1; -
DR EMBL; AEO08293; AAK89513.1; -
DR PIR; AD3037; AD3037.
DR PIR; G98248; G98248.
KM Complete proteome.
SQ SEQUENCE 431 AA; 46254 MW; 22DB764D66992C63 CRC64;

Query Match 25.0%; Score 611.5; DB 2; Length 431;
Best Local Similarity 35.5%; Pred. No. 1.2e-37;
Matches 151; Conservative 69; Mismatches 166; Indels 39; Gaps 12;

QY 34 VGTLEKEDIKFGFKITDMAPLAVALAEKGFEEDEGLFVLEAQAQWVMDRVNGELDG 93
DB 24 VQSDRQKIRAGFIPLVDASVLIAAAEFGPADREGILDLVKDVSANVRDLAFQFDI 83
QY 94 SHMLAPAPLAASVGFCTKADIEV-PSMGFNGAIVSNEIHOMKPNPL-EGKPVHP 151
DB 84 AHMLSPVPAASMLGLGNSNPSTPTTPSLRGGAITLSTRLFAFMKALTGLSETAGALEN 143
QY 152 IRADYLPVVEKKAEGK-PFNMAATFPAGSHNIKRLRYMLAAGINPGYSPQDISQI 210
DB 144 ARA--LKLVLDDMRANGAEPFLTGMYTPSSSHYERRYMLAAGIHDPH----- 190
QY 211 GADALISVTPPQMPSTLEAGTIFGVCGVBPNNQAVFKGIGVPIVITDELMKDEKVF 270
DB 191 --DVKLIVVPPMPTSDALAAAGIDGFCVGAQWNIYVAERGVRIAAKODLWPSAEKVI 248
QY 271 GYTKQMAEKYKPTTYLAVTALIRAAATWD-ADNNKRRKRAIEMLAOKQVAGDAVETLAAS 329
DB 249 GRPEWAEQOEVEGRLTALDPAASWCDLADNHDLSGA--LADPRYIQAQOSTIRRV 305
QY 330 NNGTFEYE-KDDKRALPDENTFFRHGASYSYSAVWYLTOLRRMWMINEFPDMWYLD 388
DB 306 LAGEPSIDQGNRRVYEKFTFHGDHANYPROSOSIMITYSMTKMC--QAEISGTGVNA 362
QY 389 ANVYRPDIYLAALAEKVAEKAKEDFPADTSIK-----PSONFIDKVPFANKENDYL 444
DB 363 ALSAVRPDIYRAA---LGDGKA-----PGDADIRIEQDDEGDRFVDFVPADIAQIV 413
QY 445 AKFAI 449
DB 414 NSPAV 418

```

```

NRTA_ANASP STANDARD; PRT; 440 AA.
ID NRTA_ANASP
AC Q44292; O06469;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Nitrate transport protein nrtA.
GN Name=nrtA; OrderedLocustNames=aln0608;
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
ON NCBI_TaxID=103690;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=97144534; PubMed=8990301;
RA Frias J.E., Flores E., Herrera A.;
RT "Nitrate assimilation gene cluster from the heterocyst-forming
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL J. Bacteriol. 179:477-486 (1997).
RN
RP REVISIONS.
RA Frias J.E.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
RN
RP SEQUENCE OF 1-92 FROM N.A.
RX MEDLINE=97136629; PubMed=8982006;
RA Cai Y., Wolk C.P.;
RT "Nitrogen deprivation of Anabaena sp. strain PCC 7120 elicits rapid
RT utilization of a gene cluster that is essential for uptake and
RT utilization of nitrate.";
RL J. Bacteriol. 179:258-266 (1997).
CC - FUNCTION: Essential component of the nitrate-transporting system.
CC - SUBCELLULAR LOCATION: Inner membrane-associated (Potential).
CC - SIMILARITY: Strong, to carotenoid-binding protein A (cbpA).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X99709; CA68041.2; -
DR EMBL; AP003583; BAB72566.1; -
DR EMBL; U61496; AAC46075.1; -
DR PIR; AG1882; AG1882.
DR InterPro; IPR006311; Tat.
DR TIGRFAMs; TIGR01728; Saub fam. 1.
DR TIGRFAMs; TIGR01409; Tat_signal_seq. 1.
KM Complete proteome; Inner membrane; Nitrate assimilation; Transport.
FT CONFLICT 100 W -> C (in Ref. 1).
SQ SEQUENCE 440 AA; 48475 MW; 29937A11FB45CE9C CRC64;

Query Match 25.0%; Score 610.5; DB 1; Length 440;
Best Local Similarity 33.4%; Pred. No. 1.5e-37;
Matches 158; Conservative 71; Mismatches 173; Indels 71; Gaps 17;

QY 8 SSKLLLT-----LSASLAWGLTI-----APDVGAAGV--TLEKEDLK 44
DB 5 SRKRLFTTGAAMAAASILVHGCTSGSOSATTGEOAPSAAPAAVYSAANAPVETTKAKL 64

```



```

QY 45 GFILKLTDMAPLVAAEKGFPEDEGLF-VGLEAQAAMKVMDRV---VNGELDGSHTLAP 99
DB 65 GFILPLTDAAPLILAKKGFPAKYGMTDIEVILKOKSWPVRDNKLGSSGGGIDGAHILSP 124
QY 100 APLAASVGRTKADIVPFSGMGNALITNSNEIMHMKPNIPLEGSKPHPIKADYLKP 159
DB 125 MFLVMTI--NDKVPMTILARLNTNGAISVAERF---KELNVNDE-----SSSLKD 170
QY 160 VVEKYAEKGFPMAMTFFPAGSHNKLRYWLAAGINPGYSPPODISQIGADALISVT 219
DB 171 AAIKADADKALOKGITFPGGTHDLMWRVYLAAGINP-----DDDVYLEAV 217
QY 220 PPPQMPSTLEAGTIFGVCVGEPMNQAVFKIGIVPVTDELMKDTPEKVFYTKQMAEK 219
DB 218 PPPQVAMNMKVNTVDGFCVGEPMNQAVLNOKIGSALVTELMKDHBEKAFSMRDWIEQ 277
QY 280 YPPTVYLAJVTKALIRAIWLDADNNKREKALIMLAQKQYGVADVEYLAASMGTFEYK 339
DB 278 NPMQAQIILMAILEAQOM--CDKAKENKEEMCKICSDKRYFNVAADIIERAKGNIDY--G 333
QY 340 DKRALPDF--NTFFRHGASYPSSAVWYLTOLRKMINEFKPDNWMYLDTAKNVYRPD 396
DB 334 DGKEQNFARMKFMADNMSYFYKSHDIWFLTEDIRKGYLPK---DTKQVDYVNOVWKED 390
QY 397 IYLAARKEVLAEKAKADEFPADTSIKPSQNFIDKVPDPANKRNDYLAKPAI 449
DB 391 LMKKAAR---AIGVADAE-IPASS--RGVETPF-DGVKFPDEKPEEYLSLKI 437

```

RESULT 12

Q7NEC4 PRELIMINARY; PRT; 429 AA.

```

AC Q7NEC4;
DT 01-MAR-2004 (TEMBLrel. 26, Created)
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
DE G113956 protein.
GN OrderedLocustNames=g113956;
OS Gloeobacter violaceus.
OC Bacteria: Cyanobacteria; Chroococcales; Gloeobacter.
NCBI_TaxID=33072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7421;
RX MEDLINE=22977040; PubMed=14621292;
RA Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA Sasamoto S., Matsumoto M., Kawashima K., Kishida Y., Kiyokawa C.,
RA Kohara M., Matsumoto M., Tabata S.,
RA Takeuchi C., Yamada M., Nakazaki N., Shimpō S.,
RT "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT cyanobacterium that lacks thylakoids."
RL DNA Res. 10:137-145(2003).
DR EMBL; AP006581; BAC91897.1; -.
KW Complete proteome.
SQ SEQUENCE 429 AA; 47153 MW; 205101A85C8B22P9 CRC64;

```

Query Match 25.0%; Score 610; DB 2; Length 429;

Best Local Similarity 35.4%; Pred. No. 1.6e-37; Matches 162; Conservative 60; Mismatches 174; Indels 62; Gaps 14;

```

QY 16 LSASLAWGLTI-APDVGAVGKLEKEDLKFGFIKLTDMAPLVAAEKGFPEDEGLFVQLE 74
DB 12 IAAATAITGFTLWPKYALAKGALKEPKLKYGFIALIDCAPLVIAAREGFFERBGLDVELS 71
QY 75 AQANMKVMDRVVNGELDGSHTLAPLVAASVGEFK--ADIEVPFSGMGNALITVSN 132
DB 72 KESWSMASVREGGLTGRLDASHALAGTPLAVOLGABGAPAPLITAMSLDINGNAITPSKR 131
QY 133 IMHGMKNIPLEGSKPHPIKADYLKPVVEKYAEKGFPMAMTFFPAGSHNKLRYWLA 192
DB 132 LM-----QAGVRSQ-----AD-LKKIINTGKY-GTTLTGAMVASMSMTNNLCYLAH 177
QY 193 GGINPGYSPPODISQIGADALLSVTPPPQMPSTLEAGTIFGVCVGEPMNQAVFKIG 252

```

```

DB 178 HDIHP-----YRDVRLITLTPAQILANIEAGNIDFECYTEPNTTAHAEAG 224
QY 253 VPVITDEELMKDTPPEKVFYTKQMAEKYPTVYLAATKALIRAIWLDADNNKREKALIM 312
DB 225 FVYIADRDVWGHPKEVYLAAMEPMARHTHPYTHALAKALLEACRY--CDEPARHQBVLRT 282
QY 313 LAQKQYGVADVEYLAASMGTFEYKOD-----KRALPDENTFPRHASY----- 357
DB 283 LSLRTVLAQPFILPEPALSGRLDFFGDEGKATASLKQVDPDVEVFEPKRDADYLVGNDQ 342
QY 358 --PSYSAWYLTOLRKMINEFKPD--NMYLDTAKNVPRDIYLAARKEVLAEKAKAE 414
DB 343 TFFWKSHGIMLTQMARWQIPAIIPAGVDLAD--RYRVKVDIYRQAAAEIGI--KAPAQ 397
QY 415 DFPADTSIKPSQNFIDKVPDPANKRNDYLAKPAIGLK 452
DB 398 DY-----KSENTFIDRRRFPDSNVAYLDSFEIAR 428

```

RESULT 13

Q7NKB1 PRELIMINARY; PRT; 427 AA.

```

AC Q7NKB1;
DT 01-MAR-2004 (TEMBLrel. 26, Created)
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
DE Nitrate/nitrite transport system substrate-binding protein.
GN Name=nrc4; OrderedLocustNames=g11567;
OS Gloeobacter violaceus.
OC Bacteria; Cyanobacteria; Chroococcales; Gloeobacter.
NCBI_TaxID=33072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7421;
RX MEDLINE=22977040; PubMed=14621292;
RA Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA Sasamoto S., Matsumoto M., Kawashima K., Kishida Y., Kiyokawa C.,
RA Kohara M., Matsumoto M., Tabata S.,
RA Takeuchi C., Yamada M., Nakazaki N., Shimpō S.,
RT "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT cyanobacterium that lacks thylakoids."
RL DNA Res. 10:137-145(2003).
DR EMBL; AP006573; BAC89508.1; -.
DR InterPro; IPR006311; Tat.
DR TIGRFBMS; TIGR01409; Tat_signal_seq.1.
KW Complete proteome.
SQ SEQUENCE 427 AA; 46542 MW; C0D86F1625E39273 CRC64;

```

Query Match 24.3%; Score 594; DB 2; Length 427;

Best Local Similarity 33.8%; Pred. No. 2.6e-36; Matches 152; Conservative 73; Mismatches 169; Indels 56; Gaps 14;

```

QY 9 SKLLITLSASLAWGLTIAPDVGAVGKLEKEDLKFGFIKLTDMAPLVAAEKGFPEDEG 68
DB 11 TRRGILQAAAATAATVGSILTPGYVGAADAPETTKARLGFISLSDCAPLVIAKEKGLFDKYG 70
QY 69 L-FVQLEAQAAMKVMDRV---NGLDGSHTLAPLVAASVGEFKADIEVPF-----S 119
DB 71 MKDVEVAKQASWGVTDNDELGAAGGIDGAHILTPPVYILIANGNITKSGKVPMTILAR 130
QY 120 MGFNGNAITVSNIMHMKPNIPLEGSKPHPIKADYLKPVVEKYAEKGFPMAMTFFPA 179
DB 131 LNVNGGIVSIVANK--YKL-----LKYGLDAAPKAEALTKA-----KANGDPITVAQTFPG 178
QY 180 GSHNKLRYWLAAGINPGYSPPODISQIGADALLSVTPPPQMPSTLEAGTIFGCVG 239
DB 179 GTHWAMLRVYLAAGIDP-----ETDVKAITVPPQVAMNMKVNTVDGFCV 225
QY 240 EPMNQAVFKIGIVPVTDEELMKDTPPEKVFYTKQMAEKYPTVYLAATKALIRAIWLD 289
DB 226 EPMHQDILNQDGYAVTAVTGGQIWNRRHEKSFALRADYVEKTPATKALLMAVQDAQIW-- 283

```


OY		300	ADNNKNGKEALIEMLAOKQVAGADVETLAASNMGTEFY--EKDKIKRALPFENTFRRGASV	357
Dd		284	ADKXENKQELAQVAKSRKWIGCAPSDIVARYYGIIDYGGRPVERKSPHIMQWKPAST	343
OY		358	PSYSAAVVYLTQLRWMGINEEKKPDNWYLDPAK--NVYRPDIYAARELVAEGRAKEA	414
Dd		344	PYQSHTDLWFLETEDIRMGVLPA-----TTDTKLVAANVRREDILMRBAKAL-----CG	390
OY		415	DEPADTSIKPSONEFIDKYVEPDANKPNIDL	444
Dd		391	PAPKGTG-RGVKEKF-DGVAFDPTKPEAYL	418
 RESULT_14				
ID	O8YZ76	PRELIMINARY;	PRT;	657 AA.
AC	O8YZ76;			
DT	01-MAR-2002	(TREMBlrel. 20, Created)		
DT	01-MAR-2002	(TREMBlrel. 20, Last sequence update)		
DT	01-OCT-2003	(TREMBlrel. 25, Last annotation update)		
GN	Nitrate transport ATP-binding protein.			
CN	Name=ntc; OrderedLocustNames=alr0610;			
OS	Bacteria sp. (strain PCC 7120).			
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.			
OX	NCBI_Taxid=103690;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21595285; PubMed=11759840;			
RA	Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,			
RA	Matanabe A., Iriuchimi M., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,			
RA	Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,			
RA	Yaouda M., Tabata S.;			
RT	"Complete genomic sequence of the filamentous nitrogen-fixing			
RT	cyanobacterium Anabaena sp. strain PCC 7120."			
CC	DNA Res. 8:205-213(2001).			
RL	-1- SIMILARITY: Belongs to the ABC transporter family.			
DR	EMBL; AP003583; BAB72568.1; -			
DR	PIR; A11882; A11882.			
DR	HSSP; Q58206; ILZT.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0042626; F:ATPase activity; coupled to transmembrane m. . ; IEA.			
DR	GO; GO:0015112; F:nitrate transporter activity; IEA.			
DR	GO; GO:0000166; F:nucleotide binding; IEA.			
DR	GO; GO:0015706; P:nitrate transport; IEA.			
DR	GO; GO:0006810; P:transport; IEA.			
DR	Pfam; PF00005; ABC tran.; 1.			
DR	ProDom; PD00006; ABC transporter; 1.			
DR	SMART; SMO0382; AAA; 1.			
DR	TIGRFAMs; TIGR01184; nitrCD. 1.			
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; 1.			
DR	PROSITE; PS50893; ABC_TRANSPORTER_2; 1.			
DR	ATP-binding; Complete proteome.			
SQ	SEQUENCE 657 AA; 73294 MW; F7B1434798109324 CRC64;			
 Query Match 24.0%; Score 567; DB 2; Length 657;				
Best Local Similarity 35.8%; Pred. No. 1.5e-35;				
Matches 150; Conservative 74; Mismatches 139; Indels 56; Gaps 17				
OY		39	KEDIKFGIKTLTDMAPLVAAEKGFFDEBGL-FVOLEDAQMKVKVMYMDRVNSELDSIHL	97
Dd		276	KAIVEIGMPLTDSAPLVIAKEKGFPAYGYLDNLVILNRANNQAATGVTGKLDAAOV	335
OY		98	APADLAASVGSGTADIIEV--PFSMGFNATITVENBIWHQKPINPLEGKRVHPRIKAD	155
Dd		336	AGMEIALTLTGAGSQPPTPVINAINLSRRANAATTFRKRLYNQSVRL-----AD	383
OY		156	YLKPYVEKYKAEGKPPFNAMTPPAGSHNIKLRYWLAAAGINFGYSPPODISGQIGDAL	215
Dd		384	LTKVID--SSPDQLTLTGVAHSAMQNLILRYWLAAGIID-----PRDVG-----	427
OY		216	LSVTPRPQMSESTLAGTIFGTCVGEPPNQOAVFKSIGVPVITDELMDKTPEKRGCVTKQ	275

```

Db      428  LTVIPPTOMVSQLRKGNIDGICAGPNNYQAVHDLGFYALATALEMSQPKKVLGVRED 487
Qy      276  WAEKPYNTYLAATKALIPALWLDADNNKREALEMLOKQYGVADVEVLAASNNGT- 334
Db      488  MAQKPYEYLLWALVRLIACKY--CDDLNRKREILEICRPPY----LDVNAIYVNSGFI 541
Qy      335  --EYEKD---KRALPDENFTFRHGASYPSSYSAVWYLTOLRRGMINEFKPDNNWYLDTK 390
Db      542  DPYRGGDTPPQQLTANYQFYLNKNTYNNRTEILMITYMAWG-LTFP-PKMW-VEITE 598
Qy      391  NVYRPDIYLAAKELVAEGKAKAEDPPADTSIKPSQNFIFIDKVPDPANKDNDYLAFAI 449
Db      599  RVCTDIDIGAARDL-----GLDIDGDDPI-----HLEPGKL-FNPSRPYIYLSLEI 646

```

RESULT 15	PRELIMINARY;	PRT;	420 AA.
ID 08Y2F5			
AC 08Y2F5			
DT 01-MAR-2002 (TREMBLREL. 20, Created)			
DT 01-MAR-2002 (TREMBLREL. 20, Last sequence update)			
DT 01-MAR-2002 (TREMBLREL. 20, Last annotation update)			
DE PUTATIVE NITRATE TRANSPORTER PROTEIN.			
GN Name=nasF, Synonyms=RS03349; OrderedAccession=RS03349;			
OS Ralstonia solanacearum (Pseudomonas solanacearum).			
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;			
OC Burkholderiaceae; Ralstonia.			
OX NCBI_TaxID=305;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=GM11000;			
RX MEDLINE=21681879; PubMed=11823852; DOI=10.1038/415497a;			
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mengent S.,			
RA Ariat M., Billaud A., Brottier P., Camus J.C., Cattolico L.,			
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,			
RA Gaepin C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T.,			
RA Sigulier P., Thebaud P., Whalen M., Winkler P., Levy M.,			
RA Weissenbach J., Boucher C.A.;			
RT "Genome sequence of the plant pathogen Ralstonia solanacearum. "			
RL Nature 415:497-502(2002)			
DR EMBL; AL646058; CADD3909.1; -			
DR Complete proteome.			
KM SEQUENCE 420 AA; 45516 MW; 6BC0DBA8FB82F3C CRC64;			
Query Match 24.0%; Score 585.5; DB 2; Length 420;			
Best Local Similarity 32.8%; Pred. No. 1.1e-35;			
Matches 151; Conservative 82; Mismatches 164; Indels 63; Gaps 15			
QY 3 TIIRSSSKLLITLS--ASLAVMGLTTPDVGVANGK--LEEKEDLKFGFTIKTLTDAAPLAVA 58			
DB 9 TPIPKRRRYLATVAVGASAMALIDPLVRAGAMAAAGSDAPELTBLKVGPIPLTDCASVMA 68			
QY 59 AEKGFPEDEGLFVLEAQAMWKVMDRVVNGELGSHMLAPAPLAASGVF-GTKADIEVP 117			
DB 69 ATLGDKKYGKIKIYPSKEASVAGRDCLVSGDDLAAHVLGLYGVGLIGIGPKKDMAVL 128			
QY 118 FPMGFENGAITVSNELIWHQMKPNLEBGRKVPHPIDAKDYLPVVEKKYKAEKPFNMATP 177			
DB 129 MTLNNAQOATLTSSKL-----KEAGVRDASGLKALMTTR--EKRDYTFQTF 172			
QY 178 PAGSHNKLKLYWMLAAGINPGYVSPDISOIGADLALLSTPPNPOMSTLEAGTIGYC 237			
DB 173 PTGTHAMWLYWMLAHGIRP-----IQDKAKATVPPQWVANKRVGNMDIYC 219			
QY 238 VGEPMQAAVFKGIGVITIDBELMKDTPKRVFVTKQMAEKYENTYLAATKALIRAIW 297			
DB 220 VGEPMGARALADGIGFAETQAIWKDHPEKVLGTLEAFQKYENTYRALTAALVLEASKF 279			
QY 298 LDDANNKREALEMTLAKQYGVADVVAASNMGTFEY-----EKDDRALPDNFTF 350			
DB 280 IDA-SASNRKRLTAETVAASKVYNTDMDIILDRMLGRATNGIGKTDWADDPNR-----F 331			

QY 351 FRHGA-SYSPYSAAVWYLTQLRMGMINEFKEDNNYLTDAKNVYRPDIYLAAKELVAEG 409

Db 332 YHGGVANNFPYSDGWFELTHQRKWEEL--KTHPD--YLAIAKRVNRVDIYKQA----- 381

QY 410 KAKKEDPADTSIKPSQNFIDKVPFDANKPNDIYAKFAI 449

Db 382 AATGTPPL-KSDLRTNR--LIDGVVMDAGNPPAAYADSEK 418

Search completed: May 18, 2005, 17:46:52
Job time : 184 secs

This Page Blank (uspto)